The Dissertation Committee for Daniel Turrin Fernholz
certifies that this is the approved version of the following dissertation:

Sparse Random Graphs

Methods, Structure, and Heuristics

Committee:

Vijaya Ramachandran, Supervisor

Luis Caffarelli

Adam Klivans

Greg Plaxton

David Zuckerman

Sparse Random Graphs
Methods, Structure, and Heuristics

by

Daniel Turrin Fernholz, A.B.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2007

Sparse Random Graphs

Methods, Structure, and Heuristics

Publication No. _____

Daniel Turrin Fernholz, Ph.D.
The University of Texas at Austin, 2007

Supervisor: Vijaya Ramachandran

This dissertation is an algorithmic study of sparse random graphs which are parametrized by the distribution of vertex degrees. Our contributions include: a formula for the diameter of various sparse random graphs, including the Erdős-Rényi random graphs $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$ and certain power-law graphs; a heuristic for the k-orientability problem, which performs optimally for certain classes of random graphs, again including the Erdős-Rényi models $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$; an improved lower bound for the independence ratio of random 3-regular graphs. In addition to these structural results, we also develop a technique for reasoning abstractly about random graphs by representing discrete structures topologically.

Table of Contents

Chapter 1

Introduction

A random graph is a combinatorial stranger in the strange land of probability. In some respects, combinatorics and probability are closely related — after all, combinatorial enumeration provides the foundation for discrete probability — but this kinship does not extend naturally to structural combinatorics. For this reason, the success experienced by Erdős [27] and others in analyzing discrete structures probabilistically is not only remarkable but also somewhat unexpected.

The probabilistic method represented a reversal of traditional roles: probability was used to analyze combinatorial structures, and not vice versa. Until 1959, though, random graphs remained behind the scenes, serving as means to other, non-probabilistic ends. The field of random graphs was born when Erdős and Rényi began writing about random graphs themselves, with no ulterior motives [28, 29].

Despite being both "random" and "graphs," random graphs bear only a passing resemblance to the central objects in either of these arenas. The intuitive notion of a graph as a discrete finite structure, which can be visualized as a set of "dots" connected by "lines," dates back to Euler's abstract representation of the bridges of Königsberg. Clever as it was, though, Euler's famous argument that these bridges cannot all be crossed exactly once would need substantial revision if the land masses of Königsberg were connected at random. Along with many other techniques and results from modern graph theory, Euler's proof relies heavily on precise structural knowledge, and is rendered brittle and ineffective by the uncertain nature of a random graph.

One might also conceive of a random graph as a "random object" which just happens to be a graph. However, a random graph is, after all, a discrete finite structure, which does not exhibit the algebraic properties of, say, a random variable. In an elementary sense, this

discreteness is perhaps advantageous, since there are no concerns about measurability. On the other hand, the bulk of advancements in modern probability, dating back to Kolmogorov, become somewhat superfluous, and we are back to the "balls-and-bins" enumerative style of probability.

One of the objectives of this dissertation to present what might be considered a "theory" of random graphs, rather than a disparate collection of facts. In addition to deriving various structural results about random graphs, we will develop a notion of what a graph "is": What properties can it exhibit? Which of these are important? What makes two random graphs similar or different? And, in parallel, we will develop a methodology for answering some of these questions.

Due to the interdisciplinary and eclectic nature of random graph theory, we assume no prior exposure to the field on the part of the reader. All that is assumed is a basic level of mathematical sophistication, and general familiarity with certain areas, most notably combinatorics and probability. Moreover, we will not make use of advanced concepts from either of these fields — indeed, in light of the above discussion, we would be hard pressed use many cutting-edge techniques even if we wanted to — and this dissertation is largely self-contained.

1.1 Random Graphs: The Map is Not the Territory

A random graph to can be defined as random element chosen from a given set of graphs according to a given probability distribution, but this is not what a random graph "is." This is just the map; the territory is more subtle and more elusive, and is best introduced by showing rather than telling. As an illustrative example, we shall present what is considered an important open question in random graph theory.

An *independent set* in a graph $G = (V, E)$ is a set of vertices $W \subseteq V$, no two of which are adjacent, and the *independence number* $\Gamma(G)$ is the size of a maximum independent set. For sparse graphs, this quantity is appropriately scaled by dividing by the total number of vertices $|V| = n$, and the fraction $\gamma(G) = \Gamma(G)/n$ is the *independence ratio*.

For any integers n, m, let us now denote by $\mathbf{G}_{n,m}$ a random graph chosen uniformly at random from all simple graphs with n vertices and m edges. Our problem is to find a continuous function f such that, for any $c > 0$, the independence ratio of $\mathbf{G}_{n,m}$ with $m = \lfloor cn \rfloor$ converges in probability to $f(c)$ as $n \to \infty$. Hence, for any $c > 0$, and any fixed $\epsilon > 0$, we should have

$$\lim_{n\to\infty} \mathsf{P}\left[\left|\gamma(\mathbf{G}_{n,m=\lfloor cn \rfloor}) - f(c)\right| > \epsilon\right] = 0.$$

We will not tackle this exact problem in this dissertation, but this example illustrates many of essential characteristics of the kinds of random graphs we will be analyzing, and the kinds of questions we will be seeking to answer.

1.1.1 Random Graph Models

The random graphs in the above problem are chosen uniformly among all simple graphs with n vertices and m edges. We may distinguish each actual random graph $\mathbf{G}_{n,m}$ from the idea of choosing a random graph in this way; the latter is called a *random graph model*.

$\mathbf{G}_{n,m}$ is one of two models originally proposed by Erdős and Rényi:

- $\mathbf{G}_{n,m}$ is chosen uniformly at random from all simple graphs with m edges and n vertices;
- $\mathbf{G}_{n,p}$ is a random simple graph on n vertices, in which each of the $\binom{n}{2}$ possible edges occurs independently with probability p.

What makes for a "good" random graph model is ultimately a normative question which is answered in the usual ways: elegance, simplicity, tradition, etc. For these reasons, the Erdős-Rényi models are by far the most well-studied and arguably the most important of all random graph models.

For practical purposes, these two models are equivalent; the independence ratio problem, for example, could equivalently be stated by letting $p = 2c/n$ rather than $m = \lfloor cn \rfloor$.

Despite this practical equivalence, $\mathbf{G}_{n,p}$ and $\mathbf{G}_{n,m}$ differ aesthetically in a way which is not entirely inconsequential. Indeed, an extremely strong case can be made that these are the two most natural ways to define a random graph, but choosing one over the other is more controversial. The two competing principles are uniformity of distribution and independence, and a legitimate case can be made in favor of either.

In some cases the choice of one over the other can be justified on practical terms. The edge-independence property of $\mathbf{G}_{n,p}$ facilitates deductive techniques based on independent events and random variables, while the uniformity of $\mathbf{G}_{n,m}$ makes it more amenable to arguments based on symmetry. In practice, though, symmetry arguments work almost as easily for $\mathbf{G}_{n,p}$, and near independence can be substituted for true independence to achieve the same effect for $\mathbf{G}_{n,m}$; once again, the choice boils down to aesthetic preference.

1.1.2 Asymptotics and Scaling

The independence ratio problem is not a question about just one random graph, but rather about the limit as $n \to \infty$. This is a universal principle in random graph theory: we are only interested in the asymptotic behavior as a random graph grows arbitrarily large.

In some respects, "arbitrarily largeness" is a more definitive characteristic of a random graph than its "randomness." Indeed, the classical image of dots connected by lines on a small sheet of paper might hold up, even if the seven bridges of Königsberg were chosen uniformly at random. On the other hand, if there were 10^{10} bridges, albeit deterministic ones, our visualization would have to change.

In order to deal with asymptotic sequences of random graphs, it is necessary to scale parameters appropriately, and in the above example, this is accomplished by dividing by the number of vertices. The parameter $c = m/n$ is the edge density, and since each edge is incident on two vertices, then the average vertex degree is $2c = 2m/n$. The fact that the ratio $m/n = c$ is held constant means that the random graphs in question are sparse. Sparse random graphs differ qualitatively from their denser counterparts, and in this dissertation we will deal exclusively with sparse graphs.

In addition to scaling the specification parameter $m/n = c$, the graph property in question is also scaled, in this case again by dividing the independence number by the number of vertices. This is another common characteristic of sparse random graphs: many natural properties can be expressed in terms of the size of a vertex subset exhibiting a particular behavior.

1.1.3 Convergence in Probability

The problem we posed above asks for a function $f(c)$ such that the independence ratio $\gamma(\mathbf{G}_{n,m=\lfloor cn \rfloor})$ converges in probability to $f(c)$. Implicit in this asymptotic parametrization is the belief that the structure of $\mathbf{G}_{n,m}$ somehow "converges" as $n \to \infty$, provided that the edge density is kept constant. At first glance, this may seem either obvious, surprising, or anywhere in between.

There is a certain conceptual tension here, in that a graph is a discrete structure, which is not accustomed to converging to a limit. Now, there is nothing unusual about studying discrete structures asymptotically; this is standard practice in many branches of combinatorics, and universal in theoretical computer science. It is one thing, though, to conceive of discrete structures which are arbitrarily large, and it is quite another to ask that a sequence of them converge to a limit. Despite certain recent investigations [49, 48], what exactly it means for a sequence of graphs to converge to a limit, particularly for sparse graphs, is not entirely understood.

For this particular problem, of course, we are not asking the entire random graph to converge, but only the independence ratio. However, the same kind of convergence is believed to occur for almost every other imaginable graph property, provided that the scaling is done correctly, and this amounts to a sort of central limit theorem for random graphs.

In the asymptotic limit, then, the "randomness" of a random graph is somewhat illusory. We may not know whether any given pair of vertices is connected by an edge, but this becomes largely irrelevant, first because the graph is too large for us to care about such

details, and, second, because the important properties, such as the independence ratio, are not noticeably affected by the presence or absence of a single edge.

Finally, we note that the conjectured relationship between the edge density $c = m/m$ and the independence ratio is continuous; in this case, in fact, it can be shown rather easily that if such a function f exists, it must be continuous. Once again, this continuity is somewhat at odds with the discrete nature of a graph.

Ultimately, then, what are ostensibly the three definitive characteristics of a random graph — finiteness, randomness, and discreteness — all vanish in the asymptotic limit, and what remains is something of a different nature entirely.

1.1.4 Results and Methods

We conclude this introductory example by discussing potential solutions to the independence ratio problem. For obvious reasons, it is beyond the scope of this introduction to actually give any kind of proof, so we will be brief.

For average degree $2m/n = 2c$ up to the mathematical constant e, the problem was solved by Karp and Sipser [43]. For $2c > e$, though, we find ourselves in what is a very common situation in random graph theory. There are known upper and lower bounds on $f(c)$, but the actual value is not known, and it has also not been proved that the independence ratio converges to a fixed limit.

Moreover, this author is not aware of any conjecture about the value $f(c)$. In some cases, conjectures of this sort have been arrived at using techniques from statistical physics, and it would thus not be surprising if such a conjecture were available in this particular case. However, for problems of this sort, even producing a well-motivated conjecture is a non-trivial endeavor.

As noted, it is difficult to describe how a problem like this might be solved in a short introduction, but as a starting point, we may observe that, although the expected vertex degree is $2c = 2m/n$, the actual degree of every vertex is a random variable. Moreover, in the limit, each vertex degree is Poisson distributed, and thus the fraction of vertices with

no incident edges will converge to e^{-2c}.

Since any isolated vertex will belong to every maximum independence set, we have an immediate lower bound of $f(c) \geq e^{-2c}$, albeit a trivial one. A slightly better lower bound might be achieved by then considering vertices of degree 1, of which there will be a fraction of $2ce^{-2c}$. Of course, we cannot just add them all to the independent set, since some of these may be adjacent to each other. Nevertheless, the fraction of vertices of degree 1 which are not adjacent to each other is easy to compute, and so we may achieve a stronger, but still rather trivial, lower bound.

We can continue this way, but we will not get very far, and to get beyond these naive bounds, a more coherent strategy is required. The lower bound of Karp and Sipser was achieved by tracing the execution of a greedy heuristic on the random input graph $\mathbf{G}_{n,m}$, and then proving that this heuristic is asymptotically optimal for $2c < e$. This is called the *algorithmic method*, and, aside from being quite powerful, it has the side-benefit of not only solving the independence ratio of $\mathbf{G}_{n,m}$, but also proving that this NP-hard optimization can be solved by heuristic on a random input graph.

1.2 Towards a Theory of Random Graphs

The independence ratio problem discussed above is a "good question" in random graph theory, but a collection of good questions does not a theory make. To the extent possible, we will attempt to place the results and techniques developed in this dissertation into a more general theoretical framework. Our approach is guided by two organizing principles:

1. conditional uniformity of unknown parameters,
2. topological representation of discrete structures.

The principle of conditional uniformity, simply stated, is that "what we don't know" should be uniformly random, conditional on "what we do know." The need for topological representation arises from the fact that graphs are discrete objects, but in order for something to converge to a limit, in one way or another, a topology is required.

At a nominal level, these principles are aesthetic preferences with respect to the "correct" way to talk about discrete random structures. Of the two Erdős-Rényi random graph models, for example, we would be inclined to choose $\mathbf{G}_{n,m}$ over $\mathbf{G}_{n,p}$, since $\mathbf{G}_{n,m}$ is uniformly distributed conditional on the observable parameters $|V| = n$ and $|E| = m$. And, in the example discussed above, we might describe the edge density $c = m/n$ and the independence ratio $\gamma(G) = \Gamma(G)/n$ as "topological representations" of a graph in the space $\mathbb{R}$, but this is admittedly a bit silly.

In a more abstract sense, though, the objective is to both describe and manipulate the asymptotic limit of a random graph using continuous functions. For example, if α and β are mappings from the "set" of all graphs to topological spaces X_1 and X_2, respectively, then, given a (continuous) mapping $f : X_1 \to X_2$, we can express a "theorem" about random graphs by

$$\lim_{|\mathbf{G}|\to\infty} \alpha(\mathbf{G}) = x_1 \quad \Longrightarrow \quad \lim_{|\mathbf{G}|\to\infty} \beta(\mathbf{G}) = f(x_1).$$

This would mean that, if $\mathbf{G}$ is uniformly random conditional on the parameter $\alpha(\mathbf{G}) \in X_1$, and if $\alpha(\mathbf{G}) \to x_1$ as the size $|\mathbf{G}|$ grows arbitrarily large, then the parameter $\beta(\mathbf{G})$ must converge in probability to $f(x_1)$.

Now, when dealing with a single problem involving two $\mathbb{R}$-valued parameters, such as the independence ratio, aesthetic principles and topological spaces are somewhat superfluous, but when the parameter spaces become more complicated, this abstraction becomes useful. For instance, provided that conditional uniformity is respected, it is not necessary to choose a random graph model upfront; the model is determined simply by which properties appear on the left-hand side of the implication sign. As such, "theorems" about random graphs can be chained together by composition of continuous functions.

Conditional uniformity and topological representation will play an important role in our deductive techniques, and not just in theorem statements. Once the basic machinery is put in place, the topological abstraction can also be used to state "lemmas," which, by themselves, do not yield any structural insight into a random graph, but which can be used to facilitate this sort of abstract logical inference. Ultimately, we will develop a small

library of such building-block lemmas, and most of the "theorems" in this dissertation will be proved in this more abstract setting, by manipulating topological representations and continuous mappings, and with the aid of these foundational lemmas.

1.2.1 Parametrization by Degree Distribution

As noted above, given a choice between $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$, the principle of conditional uniformity would dictate a preference for $\mathbf{G}_{n,m}$ over $\mathbf{G}_{n,p}$. However, the random graph model we ultimately settle upon will take into account not only the total number of edges, but the exact number of edges incident on each vertex. This number is the *degree* of a vertex, and the multiset of all vertex degrees can be be characterized by the *degree distribution*, which is the probability distribution corresponding to the degree of a uniformly random vertex:

$$\lambda(i) = \frac{|\{v \in V : \deg(v) = i\}|}{|V|}.$$

There are various technicalities relating to this method of asymptotic parametrization. Most notably, unlike for $\mathbf{G}_{n,m}$ or $\mathbf{G}_{n,p}$, it is not immediately clear how one would go about generating a uniformly random graph with a prescribed degree distribution. The standard approach is to use the *configuration model*, which was developed by Bóllobas [11] and Bender and Canfield [9] for this purpose, and, at a technical level, this involves working with combinatorial structures called configurations, which are similar to, but not identical to graphs.

Due to this and other minor issues, a certain amount of upfront investment is required before this parametrization by degree distribution can be managed successfully. Ultimately, though, the configuration model is more versatile than the Erdős-Rényi models; in fact, the Erdős-Rényi models can be simulated by the configuration model by specifying a Poisson degree distribution.

1.2.2 The Algorithmic Method

The *algorithmic method* refers to a collection of algorithmic techniques for studying random combinatorial structures. There are many possible variations on this central theme, but the basic idea is to predict the output of an algorithm which computes a given graph property, when the input is a random graphs. The algorithmic approach has proved quite powerful, and notable successes include lower bounds on the 3-coloring threshold for $\mathbf{G}_{n,p}$ [1], the size of a maximum matching in sparse $\mathbf{G}_{n,p}$ [43], and the size and degree distribution of the giant component of a sparse graph with fixed degree distribution [55, 56].

The algorithmic analysis of random graphs differs from the traditional analysis of algorithms in that we are primarily interested in the output of an algorithm rather than its running time. An algorithm, in this context, is not an abstract model for computation, but a way of modifying the structure of a graph in such a way that facilitates the determination of a given graph property. Given a random input graph $\mathbf{G}_0$, the execution path of such an algorithm will be a sequence of random graphs $(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_T)$, each of which is the result of some minor modifications performed on its predecessor $\pi : \mathbf{G}_t \mapsto \mathbf{G}_{t+1}$.

The algorithmic method fits naturally within the general framework of conditional uniformity and topological representation. The algorithms we analyze, in addition to computing some relevant graph property, will preserve uniformity in distribution conditional on the observable parameters. As such, the execution path can be characterized by the sequence of observables, which is the *observable process*. In our case, the observable parameter is again the degree distribution, so the observable process will be a random sequence $(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_T)$ of distributions.

Moreover, since the state only changes slightly during each step of an algorithm, by scaling the time coordinate, the observable process can be represented as an "almost continuous" function $\xi \mapsto \boldsymbol{\lambda}_{\lfloor \xi T \rfloor}$ from the interval $\xi \in [0, 1]$ to the space of distributions. In essence, the entire execution path is being represented topologically as a random element in a function space.

The increments of the underlying discrete process closely resemble the derivative of

this functional representation, and in many cases, the discrete behavior of a graph algorithm can be translated to a system of differential equations which determine the limit to which this functional representation converges in probability. The use of differential equations technique in this way, which was pioneered by Wormald [65], is one of the central tools in this dissertation.

1.3 Applications and Related Fields

This dissertation is a theoretical study of structural and algorithmic properties of sparse random graphs. As such, our driving motivation is to better understand what we consider to be important mathematical structures. Nevertheless, we are not oblivious to the potential applications of random graph theory, several of which we will now discuss.

There are two rough categories of applications. The first involves using random graphs as models for large networks that occur in the real world, either in nature or due to human activity. The second involves using random graphs to further the development of a related scientific field. In both cases, the potential for useful applications far exceeds what has been accomplished to date, and the limiting factor is not the amount of applied research effort, but the theoretical state of the art.

One issue is that current body of results about random graphs deals primarily with the Erdős-Rényi models, which only cover a very small region of the space of "possible" random graphs. While various other random graph models are available, few of these these have been studied with any real sophistication — notable exceptions include random regular graphs, random K-SAT problems, random hypergraphs, and the configuration model — and the deepest results in random graph theory are largely confined to this somewhat restricted set of random graphs.

Moreover, even for the Erdős-Rényi models, the majority of structural properties are not well-understood. As such, for the foreseeable future, progress in applications of random graph theory is likely to remain closely correlated to progress on the theoretical front.

1.3.1 Random Graphs in the Real World

The idea of using graphs to model real-world phenomena is not new; indeed, the "first" graph was an abstract representation of land masses connected by bridges. In the last decade, though, there has been an explosion of interest in studying the structure of various real-world networks, which unlike the Königsberg bridge graph, resemble random graphs more than their discrete counterparts.

The prototypical examples of such real-world networks are the internet and the web graph (the graph of web pages and hyperlinks). Whether or not these are truly "random" is perhaps a philosophical question, but they are clearly the product of many small, independent modifications, rather than a single grand design. Moreover, these graphs are very large, even "massive," as they are often described. And, as with random graphs, their exact structure is not precisely knowable, and generally not relevant; if a single link were missing from the web graph, nothing much would change.

Massive real-world graphs, however, do not often resemble Erdős-Rényi random graphs. Most notably, rather than a Poisson degree distribution, real-world graphs typically exhibit so called "power-law" or "thick-tailed" degree distributions, meaning that the fraction $\lambda(i)$ of vertices of degree i drops polynomially, rather than exponentially with i. Both the internet and the web graph were found to exhibit power-law degree distributions by a number of researchers (e.g. [30, 47, 40]), and power-laws were subsequently observed in many other massive graphs including social networks, disease transmissions, and protein interactions.

The "power-law" phenomenon, which goes by many names, is nothing new: the economist's power-law is the Pareto distribution, which describes the distribution of wealth; the linguist's is Zipf's law, and is related to frequency of word usage. Even power-law graphs date back at least to 1965, when de Solla Price observed a power-law in the graph of scientific citations [24].

The source of the power-law is what de Solla Price called *cumulative advantage*, and the idea is that the rich get richer. A power-law graph degree distribution can be generated

by adding edges to a graph one at a time, and, rather than connecting two vertices uniformly at random, choosing vertices in proportion to their current degree.

The discovery (or re-discovery) of power-law graphs led to a slew of random graph models which attempted to capture the behavior of these real-world graphs, typically using one of two basic strategies. The first is to generate graphs incrementally, as described above, using the principle of cumulative advantage [7, 8] (which now goes by the name of *preferential attachment*). The second is to use the configuration model to choose uniformly at random from all graphs with a power-law degree distribution [2].

There are various problems with either approach, the most significant of which is that real-world graphs exhibit dense clusters and short cycles, while these random graphs are locally tree-like. This observation prompted another generation of models which attempt to also account for the clustering in real-world graphs. However, these second-generation models are also somewhat problematic, in that the clustering that occurs in real-world graphs is not entirely understood, and there is no satisfactory way to determine whether a given model produces the correct kind of clustering.

Despite this recent progress, the science of massive real-world networks is truly in its infancy, and the problem is not a lack of power-law models, but a generally inadequate theoretical understanding. For example, the minimal requirements for a rigorous empirical science would presumably include the ability to make predictions and test hypotheses. In other words, one ought to be able to observe certain phenomena in a real-world graph, make theoretical predictions, and then go back and see whether these predictions are accurate. This would ideally lead to some rigorous way to establish that a given real-world graph actually resembles a theoretical model. At present, the theory of random graphs is not capable of supporting such applications with any degree of sophistication.

1.3.2 Random Graphs and Computer Science

In terms of sheer volume of publications, the bulk of interest in random graphs by computer scientists is related to studying the internet and the world-wide web using random

models. However, this is not so much a computational application of random graph theory as it is an instance of a real-world graph which also happens to be of interest to computer scientists. Although understanding these networks is certainly an important objective, there are various other applications of random graph theory more substantively related to theoretical computer science.

The Algorithmic Theory of Random Graphs

This dissertation is mainly concerned with using algorithmic methods to determine structural properties of random graphs. A related research area, sometimes termed the *algorithmic theory of random graphs*, involves determining whether certain computational problems which are difficult to solve in general (i.e. NP-hard problems) can be solved efficiently on random graphs [35]. Of course, if we are able to determine a structural property of a random graph algorithmically, then de facto we have also solved the related algorithmic problem as well.

Now, if it were possible to establish rigorously that certain real-world graphs truly resembled random graph models, then the algorithmic theory of random graphs would yield rigorous results about the running time of algorithms on real-world input graphs. However, pending this development, in most settings, the assumption that the input to an algorithm is uniformly random is unrealistically strong.

At a less rigorous level, though, both the structural and algorithmic analysis of random graphs can contribute to the general understanding of what makes a problem computationally difficult. Indeed, the most difficult graph properties to solve structurally for sparse $\mathbf{G}_{n,p}$ generally correspond to NP-hard optimization problems; notable examples include the independent set problem described above, and also the chromatic number. This is partly due to the fact that, by and large, most combinatorial optimization problems are both NP-hard computationally, and also hard to solve structurally for random graphs.

However, there are at least two examples of problems which are neither NP-hard nor entirely computationally trivial, but which have been solved for sparse $\mathbf{G}_{n,p}$. The first is the

size of a maximum matching, which was solved by Karp and Sipser [43], and the second is the k-orientability problem which is discussed in chapter 10 of this dissertation. Moreover, both maximum matching and k-orientability are solvable on sparse $\mathbf{G}_{n,p}$ using relatively simple heuristics; on the other hand, empirical studies and non-trivial conjectures suggest that many NP-complete problems retain some of their computational difficulty, even if the input graph is random.

Since most NP-complete problems remain unsolved both structurally and algorithmically on sparse random graphs, the rigorous differentiation between algorithmic and structural properties of sparse random graphs is currently somewhat beyond the state of the art. Nevertheless, the limits of algorithmic methods on sparse random graphs are likely to provide for fascinating research questions in the future.

Heuristics

Regardless of whether the algorithmic theory of random graphs bears any relation to the general theory of computational complexity, a practical benefit of this algorithmic analysis is that it often leads to the development of efficient heuristics.

An example of this is the simulated annealing heuristic, which is not a direct offspring of the theory of random graphs, but which grew out of a closely related area in statistical physics. More recently, this same line of thought has led to the development of a heuristic for K-SAT and coloring problems called *survey propagation* [16] which appears promising both in theory and in experiment.

It is difficult to make rigorous claims about the performance of any heuristic; this is almost by definition. Indeed, a proof that a given heuristic performs well on uniformly random input is technically only relevant if it is accompanied by some sort of assurance that the graphs we intend to run it on are also uniformly random. Moreover, the strongest heuristics, including survey propagation, are typically not provably effective even on random input, since the insight which leads to the development of a heuristic typically precedes this kind of rigorous analysis.

Nevertheless, the true test of a heuristic is its success on real-world input, and in practice, good heuristics are enormously valuable. As such, the fact that our structural and algorithmic understanding of random combinatorial objects often leads to the development of new heuristics cannot be overlooked. In terms of practical impact, simulated annealing and other heuristics from the same family are perhaps the most significant computational spinoffs from this entire line of research.

Random Graphs as Algorithmic Tools

A third application of random graph theory to algorithm design involves actually using a random graph as a tool to achieve other algorithmic objectives. In other words, if we are able to generate a random graph (pseudo-randomness issues notwithstanding), we may then exploit known structural properties of this random graph to design efficient randomized algorithms.

An example of this appears in chapter 10 of this dissertation. The random graph property in this case is k-orientability, and the algorithmic applications are to hashing and load-balancing. If the Erdős -Renyi random graph $\mathbf{G}_{n,m}$ is k-orientable for a given edge-density $c = m/n$, then a certain generalization of Cuckoo hashing [58] can be shown to succeed with high probability, and it thus becomes critical to determine the maximum edge-density for which $\mathbf{G}_{n,m}$ is k-orientable with high probability.

1.3.3 Random Graphs and Statistical Physics

Although the field of random graphs is technically classified as a mathematical discipline, certain branches of statistical physics have been dealing with very similar structures for quite some time. In fact, due largely to the power-law phenomenon, a number of statistical physicists, along with scientists from various other disciplines, have taken a more direct interest in random graphs.

However, this is a relatively recent development, and a much deeper connection between random combinatorial structures and statistical physics can be found in the field of

spin glasses [53]. Briefly, a spin glass is a physical phenomenon which behaves very much like a random instance of a combinatorial optimization problem; a more accurate definition is beyond the scope of our present discussion.

The physicist's view of random combinatorial structures differs somewhat from that of a combinatorialist or a probabilist, and this view is accompanied by a different, and quite powerful, set of analytic techniques. In many cases, spin glass theory can only offer conjectures rather than rigorous proofs, although recently some remarkable progress has been made in making some of these conjectures rigorous [63]. Nevertheless, these conjectures are "probably" correct — in many cases, they have been verified experimentally — and the often offer a glimpse into the future of the rigorous mathematical theory of random graphs.

1.4 Organization and Contributions of this Dissertation

This dissertation is organized in four parts, followed by a brief conclusion.

Part I: Preliminaries

The first part deals with basic definitions: in chapter 2 we present some notation and review background material, and in chapter 3, we introduce the configuration model.

Part II: Methods

Our methodology is developed in chapters 4 and 5. In chapter 4, we present a somewhat abstract approach to reasoning asymptotically about discrete random structures using topological representation, including an independent proof of a version of the differential equations theorem of Wormald [65]. In chapter 5, we deal specifically with the kinds of random structures which we will be analyzing for the remainder of the dissertation: discrete random processes which arise from the algorithmic analysis of the configuration model.

Part III: Structural Properties

In the third part, we examine basic structural properties of random graphs: the decomposition into connected components, the cores, and the diameter. Our first applications appear in chapter 6, in which we analyze the 2-core of a random graph, along with its decomposition into connected components.

In chapters 7 and chapter 8, we examine the local structure of a random graph in the vicinity of a particular vertex. Chapter 7 explores the relationship between the local structure of a random graph and a branching process at a somewhat informal level. In chapter 8, we analyze the local structure more rigorously, and in doing so we compute the diameter of a sparse random graph.

In chapter 9, we generalize the analysis of the 2-core and solve the analogous k-core problem.

Part IV: Heuristics

In the final part, we examine the performance of heuristics for two combinatorial optimization problem. In chapter 10, we present a heuristic for the k-orientability problem, and we show that its performance is asymptotically optimal for a certain class of random graphs which includes the Erdős-Rényi graphs $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$; the k-orientability problem has special significance for Erdős-Rényi random graphs due to algorithmic applications to hashing and load balancing.

In chapter 11, we consider the independence ratio problem for the special case of a random 3-regular graph; we first reproduce the algorithmic analysis of the greedy heuristic by Frieze and Suen [36], and we then present a more powerful heuristic that attains a stronger lower bound.

1.4.1 Summary of Original Contributions

The original contributions of this dissertation include structural results regarding the following four random graph properties.

1. The size and degree distribution of the k-core of a random graph parametrized by degree distribution is presented in chapters 6 and 9, and appeared in [32]. For the Erdős-Rényi models $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$, the k-core threshold was found by Pittel, Spencer, and Wormald [59]. The k-core problem for an arbitrary degree distribution was solved independently by Molloy [54], Cooper [22], and Janson and Luczak [38] by similar techniques.

2. An asymptotically accurate formula, of the form $c \ln n \pm o(\ln n)$, for the diameter of a sparse random graph is presented in chapter 8, and appeared in [33]. For both sparse $\mathbf{G}_{n,p}$ and arbitrary distributions, this is the first result with this degree of accuracy.

3. The k-orientability thresholds for $\mathbf{G}_{n,p}$, along with certain other degree distributions, is presented in chapter 10, and appeared in [34] and [31]. The same result for $\mathbf{G}_{n,p}$ was established independently, using a different technique, by Cain, Sanders, and Wormald [18].

4. An improvement on the lower bound of Frieze and Suen [36] for the independence ratio of random 3-regular graphs appears in chapter 11. This result has not yet been written up as an individual paper.

We also include an alternate derivation of two significant results in random graph theory. First, a version of the differential equations theorem of Wormald [65] is proved in chapter 4. Second, an original presentation of the main results of Molloy and Reed [55, 56] regarding the decomposition of a graph into connected components is given in chapter 6.

The methodology developed in chapters 4 and 5 for reasoning asymptotically about discrete random structures is either an original contribution or it is not. This methodology does not require any "results," and is based almost entirely on topological properties of function spaces, and very basic ideas from calculus. And, each individual part of this methodology can be traced to another origin:

- the method of parametrizing the configuration model using convergence of the degree distribution is due to Molloy and Reed [55, 56];

- the use of differential equations for solving discrete random processes is due to Wormald [65];
- the idea of topological representation is ubiquitous in mathematics, and has been discussed in the specific context of graph theory in several ways [49, 48];
- our general methods of asymptotic reasoning draw heavily upon a number of classical ideas, including non-standard analysis, lattice completions, and representation theorems for Boolean algebras.

On the other hand, the whole in this case is perhaps greater than the sum of its parts, and bringing these elements together under one roof yields a coherent methodology which is intended for use beyond the specific applications in this dissertation.

Part I

Preliminaries

Chapter 2

Notation and Background Material

This dissertation is largely self-contained, and assumes no prior exposure to the field of random graphs on the part of the reader. The only real technical prerequisite is a general familiarity with the following four areas:

1. graph theory and combinatorics,
2. probability,
3. point-set topology,
4. real analysis and differential equations.

The purpose if this chapter is to establish basic notation and conventions, and to present some preliminary background results. It is not intended as an introduction to any of the four areas described above. For more background, we refer the reader to any number of text on these topics, including: for graph theory [14], and for random graphs in particular [13] and [39]; for probability [10] and [42]; for topology [57]; for real analysis [10] and [60]; for differential equations [21].

Chapter Organization

In section 2.1, we give an overview our notational conventions. In section 2.2, we review some relationship between pointwise and ℓ_1 convergence in the space $\mathbb{R}^\omega$. In section 2.3, we review Chernoff's large deviations inequality, and present some corollaries.

2.1 Basic Notation and Conventions

The field of random graphs, by nature, gives rise to a certain amount of notation. First, we have the basic problem of describing combinatorial structures algebraically. Second, there is notation related to probabilities, expectations, and so on. Third, random graphs are studied asymptotically, which introduces another layer of notation. In our case, the notational complexity is compounded by the fact that our primary technique is the algorithmic method, which involves random processes, rather than individual random objects.

In general, we shall abide by established mathematical convention to the extent possible. However, in order to alleviate some of the notational burden, we will adopt a number of simplifying abbreviations and conventions which are not entirely standard. The most important points are summarized below, and each topic is discussed individually in its own subsection.

§2.1.1: Random Structures and Discrete Probability

- Random variables and structures are denoted in **boldface**.
- We are working with discrete probabilities, so there are no concerns about measurability.
- A *generic* random object is an equivalence class modulo equivalence in distribution.
- A *random mapping* $\mathbf{f} : S \to R$ maps S to the set of generic random elements in R, so $\mathbf{f}(s)$ is a random element in the set R.

§2.1.2: Asymptotics

- An *asymptotic sequence* is a sequence which exists for the sole purpose of taking limits.
- The symbol η is reserved exclusively to index asymptotic sequences, as in s_η.

- An *asymptotic element* in S is implicitly assumed to be an asymptotic sequence, and is written simply as "s" without the subscript.
- Unqualified references to "convergence," uses of the symbol "$\to$," or "big O" notation to the limit as $\eta \to \infty$.
- An *asymptotic random element* is an asymptotic sequence of random elements, and in this case convergence means convergence in probability.

§2.1.3: Graphs and Configurations

- We will technically be dealing with combinatorial structures called *configurations*, which are defined in the next chapter.
- Configurations and graphs are similar, and for general discussion, we shall use the term "graph" to refer to either of these, use the same symbol G, etc.

§2.1.4: Naive Set Theory

- A function, or a mapping, from $R \to S$ is more or less the same an element in the cartesian product S^R and we may use either kind of notation:
  - a "function" is written $f : R \to S$ as usual,
  - a "tuple" is written $(s_r)_{r \in R}$ and abbreviated by (s_r).
- A function $f : R \to S$ acts naturally on subsets by $f(R_0) = \{f(r_0) : r_0 \in R_0\} \subseteq S$.
- We will not run into any set-theoretic dilemmas:
  - a *universal set* is a "set of all" objects of a given kind,
  - adequate universal sets can generally be constructed rigorously, and we shall denote these using blackboard fonts (e.g. $\mathbb{G}$ denotes the set of all configurations),
  - choice functions exist if we need them to, and this does not generally require the axiom of choice, not that it would matter if it did.

§2.1.5: Product and Function Spaces

- $\mathbb{R}^\omega$ is a countable product of $\mathbb{R}$:
  - the countable index set is assumed to be $\mathbb{Z}^* = \{0, 1, \ldots\}$,
  - elements in $\mathbb{R}^\omega$ are generally written in "function" notation by

$$x = (x(0), x(1), \ldots).$$

- $\mathcal{M}([0,1], S)$ is the set of functions from $[0,1]$ to S:
  - elements in $\mathcal{M}([0,1], S)$ are generally written in "tuple" notation by

$$(s_\xi) = (s_\xi)_{\xi \in [0,1]}.$$

- $\mathcal{C}([0,1], X)$ is the set of continuous functions from $[0,1]$ to a topological space X.
- All products of topological or metric spaces endowed with the product topology by default; the other topologies we use are
  - $x_\eta \xrightarrow{\ell_1} y$ denotes ℓ_1 convergence in $\mathbb{R}^\omega$;
  - $(x_{\xi,\eta}) \xrightarrow{\infty} (y_\xi)$ denotes uniform convergence in $\mathcal{M}([0,1], X)$, for any metric space X.

2.1.1 Random Structures and Discrete Probability

Our notation and conventions for random objects can be briefly summarized by stating that we intend to treat random elements in a set S in the same way as an ordinary elements, except that they are "random." We shall use boldface to distinguish random objects, and this allows us to introduce a random element by simply writing $\mathbf{s} \in S$; since $\mathbf{s}$ does not actually belong to the set S, this is an abuse of notation, but the meaning is clear.

Since we are studying discrete random structures (i.e. graphs), we may assume that all probabilities are discrete which means, essentially, that any random element $\mathbf{s} \in S$ has countable support, in the sense that there is some countable set $S_0 \subseteq S$ such that $\mathsf{P}[\mathbf{s} \in S_0] = 1$, and $\mathsf{P}[\mathbf{s} = s] > 0$ for each $s \in S_0$. This allows us to define random

elements in any set, without needing to first construct a measure. It also allows us to ignore issues related to sets of measure zero; we shall generally use the terms "always" and "with probability 1" synonymously.

Distributions

Discrete distributions are described using the following notation:

- the *distribution* of a random element $\mathbf{s} \in S$, is the mapping $S \mapsto [0,1]$ given by $s \mapsto \mathsf{P}[\mathbf{s} = s]$.
- $\mathfrak{D}[\mathbf{s}]$ denotes the distribution of $\mathbf{s}$, so $\mathfrak{D}[\mathbf{s}](s) = \mathsf{P}[\mathbf{s} = s]$,
- $\mathrm{Dist}(S)$ denotes the set of all (discrete) distributions on S,
- $\mathfrak{U}(S)$ denotes the uniform distribution on a finite set S,
- $\mathbf{s}_1 \stackrel{d}{=} \mathbf{s}_2$ denotes equality in distribution meaning that $\mathfrak{D}[\mathbf{s}_1] = \mathfrak{D}[\mathbf{s}_2]$.

Generic Random Objects and Random Mappings

We shall generally avoid explicit references to any underlying probability space by using what we shall call *generic* random objects. Technically, we define a generic random element in a set S to be an equivalence class modulo equivalence in distribution. Intuitively, a generic random element $\mathbf{s}$ is more or less an algebraic device for manipulating a distribution using "probabilistic" notation.

A second important concept is that of a *random mapping*, which is a mapping from a set S to the set of generic random elements in a second set R. Similarly to random elements, we may introduce a random mapping by writing $: \mathbf{S} \to \mathbf{R}$, which is technically an abuse of notation, but again the meaning is clear. Intuitively, a random mapping can be understood as a "black box," which takes as input a random element from S, and outputs an element in R which is "random."

We have the following notation and conventions:

- $\mathbf{x}_\mu$ denotes a generic random variable with distribution μ;
- generic random objects are independent by default, so $\sum_{i=1}^{j} \mathbf{x}_{\mu,i}$ denotes the sum of j i.i.d. copies of $\mathbf{x}_\mu$;
- each call to a random mapping is similarly independent, and thus a Markov chain can be specified using a random mapping $\mathbf{f} : S \to S$ by letting $\mathbf{s}_{t+1} = \mathbf{f}(\mathbf{s}_t)$ for all t.

Since we will omit references to an underlying probability space throughout this dissertation, then every random object is considered generic, and we will shortly stop explicitly using the term "generic."

Now, the fact that generic random objects are independent by default does not mean that they are always independent. For example, given random variables $\mathbf{x}, \mathbf{y}$, we may specify that

$$\mathbf{x} = \mathbf{y} + 1.$$

This means that the pair $(\mathbf{x}, \mathbf{y})$ is a generic random element in $\mathbb{R}^2$ which satisfies $\mathbf{x} = \mathbf{y} + 1$ always (or, equivalently, with probability 1). Note that this is not the same as the condition $\mathbf{x} \stackrel{d}{=} \mathbf{y} + 1$.

Conditional Probabilities and Distributions

For random elements $\mathbf{s}, \mathbf{r} \in S, R$ and an event L:

- $\mathfrak{D}[\mathbf{s} \mid L]$ is the distribution of $\mathbf{s}$, conditional on L;
- $\mathsf{P}[L \mid \mathbf{s}]$ is the probability of L, conditional on $\mathbf{s}$;
- $\mathfrak{D}[\mathbf{r} \mid \mathbf{s}]$ is the distribution of $\mathbf{r}$, conditional on $\mathbf{s}$.

Note that, while $\mathfrak{D}[\mathbf{s} \mid L]$ is just an ordinary distribution on the set S (provided that $\mathsf{P}[L] > 0$), both $\mathsf{P}[L \mid \mathbf{s}]$ and $\mathfrak{D}[\mathbf{r} \mid \mathbf{s}]$ are random objects, since they depend on the value of $\mathbf{s}$. In particular, $\mathfrak{D}[\mathbf{r} \mid \mathbf{s}]$ is a random element in the set $\mathrm{Dist}(R)$ of distributions on R, and in our discrete setting, there is no question about whether or not this random element is well-defined.

Random Processes

We shall use "tuple" notation to denote random processes, so an S-valued random process would be denoted by

$$(\mathbf{s}_t)_{t=0}^{T} = (\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_T),$$

and may be abbreviated by simply $(\mathbf{s}_t)$.

The time coordinate is by default assumed either to range from $t = 0$ to either ∞ or some fixed time T. We call T the *specified duration* of the process. In some cases, the process may halt prematurely at some random time $\boldsymbol{\tau} \in \{0, \ldots, T\}$; in this case, we abide by the convention that the sample path is extended for the specified duration in the halted state, so $\mathbf{s}_t = \mathbf{s}_\tau$ for $t > \boldsymbol{\tau}$.

For the purposes of taking conditional probabilities and expectations, we will generally consider the "state of knowledge" at time t to be the history, so the conditional expectation "at time t" is by default assumed to mean

$$\mathsf{E}_t[\cdot] = \mathsf{E}[\cdot \mid \mathbf{s}_0, \ldots, \mathbf{s}_t].$$

Conditional probabilities $\mathsf{P}_t[\cdot]$ and conditional distributions $\mathfrak{D}_t[\cdot]$ are defined in the same way.

This is not a universal rule though, and in some cases, we will include other state information in addition to the history, but in these cases, we will do so explicitly.

Increments and Martingales

For an $\mathbb{R}$-valued process $(\mathbf{x}_t)$:

- $\Delta\mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$ denotes the increment of an $\mathbb{R}$-valued process $(\mathbf{x}_t)$ at time t,
- a *martingale* is a process such that $\mathsf{E}_t[\mathbf{x}_{t+1}] = \mathbf{x}_t$, or equivalently, $\mathsf{E}_t[\Delta\mathbf{x}_{t+1} = 0]$, for all t.

2.1.2 Asymptotics

An *asymptotic sequence* is a sequence which exists for the sole purpose of taking limits. We shall reserve the symbol η exclusively to index asymptotic sequences, as in s_η, and in most cases we will leave of this symbol and take the limit $\eta \to \infty$ implicitly. Accordingly, an *asymptotic element* in S is an asymptotic sequence which is written simply as "s" rather than "s_η."

When working with asymptotic elements, any unqualified references to "convergence" refer to the limit along this sequence; for instance, if x is an asymptotic element in $\mathbb{R}$, then $x \to 0$ means that $\lim_{\eta\to\infty} x_\eta = 0$. This of course does not mean that other convergences cannot take place, only that they must be qualified; for instance, the statement "$f(z) \to 0$ as $z \to 0$" indicates an ordinary convergence which has nothing to do with asymptotic sequences.

Asymptotic Notation

We shall frequently use the standard "big O" notation:

- $x_\eta = O(y_\eta)$ means that there exist constants $C, N > 0$ such that $x_\eta < Cy_\eta$ for $n > N$;
- $x_\eta = o(y_\eta)$ means that, for all $\epsilon > 0$, there exists N such that $x_\eta < \epsilon y_\eta$ for $n > N$;
- $x_\eta = \Omega(y_\eta)$ means that $y_\eta = O(x_\eta)$;
- $x_\eta = \omega(y_\eta)$ means that $y_\eta = o(x_\eta)$;
- $x_\eta = \Theta(y_\eta)$ means that both $x_\eta = O(y_\eta)$ and $x_\eta = \Omega(y_\eta)$.

As usual, unqualified uses of "big O" notation refer to the limit as $\eta \to \infty$.

Convergence and Separation

We will often work with the topological definition of convergence; when working with a topological space X, an expression of the form "$B \ni x$" for some $x \in X$ is understood to indicate that B is an open neighborhood of x in X.

For an asymptotic sequence x_η in a topological space X, and a fixed $y \in X$:

- $x_\eta \to y$ means that, for all $B \ni y$, there exists N such that $x_\eta \in B$ for all $\eta > N$;
- $x_\eta \nrightarrow y$ means that there exist $B \ni y$ and N such that $x_\eta \notin B$ for all $\eta > N$.

The first of these definitions is of course standard, but second is not. We shall call the condition $x_\eta \nrightarrow y$ a *separation*, and the symbol "$\nrightarrow$" should be interpreted as "is bounded away from" rather than "does not converge to."

Asymptotic Random Elements

An *asymptotic random element* in a set S is an asymptotic sequence $\mathbf{s}_\eta$ or random elements in S. Asymptotic random elements will be discussed at length in the next two chapters, and the above definitions will be generalized to deal with convergence in probability of asymptotic random elements.

For now, we only point out that an asymptotic random element is not an S-valued random process, but rather an asymptotic sequence of unrelated random elements, and in particular, any limits as $\eta \to \infty$ are limits in probability.

2.1.3 Graphs and Combinatorial Structures

The terminology used to describe graphs and related combinatorial structures has not been entirely standardized. At a technical level, we will be working with configurations (which are discussed at length in the next chapter) rather than graphs, so we will use graph-theoretic terminology primarily for high-level, informal discussions, and any ambiguities will be resolved at the time they arise.

Basics

A *graph* is a pair $G = (V, E)$ where V is a set of *vertices*, E is a set of *edges*, and each edge is an unordered pair of (not necessarily distinct) vertices $e = \{u, v\}$. The following definitions are relatively non-controversial.

- an edge $e = \{u, v\}$ is said to *connect* the vertices u, v;
- in this case, u and v are said to be *adjacent*, and the edge e is *incident* upon both u and v;

Simple Graphs

A graph, as we have defined it, allows for loops and parallel edges, which are defined as follows:

- an edge $e = \{v, v\}$ which connects a vertex v to itself is called a *loop*;
- a set of two or more edges $e_1 = \{u, v\}$ and $e_2 = \{u, v\}$ which connect the same pair of vertices is called a set of *parallel edges*;
- a graph $G = (V, E)$ is *simple* if the edge set E does not contain any loops or parallel edges;

The term *multi-graph* is often used to describe a graph which is not necessarily simple. For our purposes, though, the terms "graph" and "multi-graph" will be considered synonymous, so by default a graph G may contain loops or multiple edges unless we specify that G is simple.

Vertex Degrees

An edge $e = \{u, v\}$ is said to be *incident* upon vertices u and v, and the *degree* of a vertex v is the number of edges incident upon v, counting multiplicities, so a loop $\{v, v\}$ is counted twice. A vertex is *isolated* if its degree is 0.

Isomorphisms

An *isomorphism* between graphs $G = (V, E)$ and $G = (V', E')$ is a pair of bijections $f : V \to V'$ and $g : E \to E'$ such that, for every $e = \{u, v\}$, $g(e) = \{f(u), f(v)\}$. And, two graphs are *isomorphic* if there exists an isomorphism between them. Similarly, for any pair

of of combinatorial objects of any kind, an isomorphism is a set of bijections between the various components which preserves all structural relations in this way.

2.1.4 Naive Set Theory

We will not tangle with any foundational issues in set theory. Our present purpose is to clarify some notation, and show how *universal sets* of "all" objects of a given kind can be constructed easily and without controversy.

Functions, Mappings, and "Tuples"

The words "function" and "mapping" are synonymous for our purposes, and a "tuple" is simply a mapping (or function) which is written $r \mapsto s_r$ rather than $r \mapsto f(r)$. The choice of whether to use "function" versus "tuple" notation is more or less arbitrary, and depends upon how we want to use the particular object.

With at least two important exceptions (see §2.1.5), we will denote by S^R the space of such mappings, and generally use "tuple" notation to denote elements of S^R. We denote an entire tuple by $(s_r)_{r\in R}$, and abbreviate by (s_r); the expression s_r alone denotes a single element in S corresponding to the coordinate r of this tuple.

Some related notation:

- $(s_i)_{i=1}^t$ denotes the sequence $(s_1, s_2, \ldots, s_t)$,
- $\{s_r\}_{r\in R}$ denotes the *multiset* of all s_r (including multiplicities) for $r \in R$,
- $\{s_r : r \in R\}$ denotes the corresponding subset of S (without multiplicities), which can also be written $\bigcup_{r\in R}\{s_r\}$.

We will often abuse notation by letting a function $f : R \to S$ act on subsets as follows:

- for a subset $R_0 \subseteq R$, we let $f(R_0) = \{f(r_0) : r_0 \in R_0\} \subseteq S$;
- for a subset $S_0 \subseteq S$, we let $f^{-1}(S_0) = \{r \in R : f(r) \in S_0\} \subseteq R$.

Less frequently, we will abuse notation in the opposite way, by identifying any $s \in S$ with the singleton subset $\{s\} \subseteq S$, as in $f^{-1}(s) = f^{-1}(\{s\}) = \{r \in R : f(r) = s\}$.

Given a subset $S_0 \subseteq S$, the set difference is denoted by

$$S \backslash S_0 = \{s \in S : s \notin S_0\}.$$

Universal Sets

We will often find it useful to discuss universal sets such as the "set of all (finite) graphs." This has nothing to do with formal set theory, and is mainly to describe some simple ideas in terms of sets and mappings.

As an example, we will show how a suitable set of all graphs might be defined. Since $\mathbb{G}$ will denote the set of all configurations, we shall use the symbol $\hat{\mathbb{G}}$ to denote the set of all graphs. In this case, we may first designate a set of all "vertices," which we may denote by $\mathbb{V}$; a good choice would be $\mathbb{V} = \mathbb{N}$. Then, we let $\hat{\mathbb{G}}$ denote the set of all graphs for which the vertex set is a (finite) subset of $\mathbb{V}$.

We may now, e.g., define a "graph property" as a mapping $\alpha : \hat{\mathbb{G}} \to \{\top, \bot\}$. Note, though, that we are not imposing any canonical form on the set of all graphs. This allows us to, e.g., place two distinct copies of the same graph next to each other; also, a sub-graph of a graph in $\hat{\mathbb{G}}$ is also an element of $\hat{\mathbb{G}}$, and so on.

Choice Functions

Along with universal sets, we will also occasionally use choice functions, primarily for use in algorithmic constructions. The idea is that an algorithm often needs to "choose an arbitrary vertex" in a graph, and we may describe this operation set-theoretically in terms of a choice function $v_{\mathcal{C}}$ from the set $\hat{\mathbb{G}}$ to the set $\mathbb{V}$ of vertices, such that $v_{\mathcal{C}}(V, E) \in V$ for every graph (V, E).

Again, this is neither a deep nor a controversial idea; if the universal vertex set is $\mathbb{V} = \mathbb{N}$, then we may explicitly construct such a choice function by choosing the minimal $v \in V$, and there is no need for the axiom of choice.

2.1.5 Product and Function Spaces

There are two specific kinds of product and function spaces which will play a special role in this dissertation.

The Countable Product $\mathbb{R}^\omega$

$\mathbb{R}^\omega$ denotes a countable cartesian product of $\mathbb{R}$, and although elements of $\mathbb{R}^\omega$ are arguably "tuples," we will denote them as "functions." In a second departure from convention, we will use the set $\mathbb{Z}^* = \{0, 1, \dots, \}$ as the countable index set, so an element in $\mathbb{R}^\omega$ is written

$$x = (x(0), x(1), \dots).$$

The reason for doing this, is that countable product spaces will most frequently be used to describe probability distributions on the set $\mathbb{Z}^*$, which will in turn be used to parametrize our random graphs. We shall denote the set of such distributions by

$$\Phi = \mathrm{Dist}(\mathbb{Z}^*),$$

and we note that Φ is a subset of $[0,1]^\omega$, which is in turn a subset of $\mathbb{R}^\omega$.

Elements of Φ will typically be written using the Greek letters λ, μ, and ν. By using "function" notation, we are able to denote an element of Φ by a single symbol, rather than a "tuple."

The Function Spaces $\mathcal{M}([0,1], S)$ and $\mathcal{C}([0,1], X)$

For an arbitrary set S and a topological space X:

- $\mathcal{M}([0,1], S)$ is the set of functions $[0,1] \to S$;
- $\mathcal{C}([0,1], X)$ is the set of continuous functions $[0,1] \to X$.

Although elements of either of these sets are "functions," we will use "tuple" notation:

$$(s_\xi) = (s_\xi)_{\xi \in [0,1]}.$$

The reason is that these objects will be used to represent the scale sample path of a random process. This is described in chapter 4, but briefly, for a random process $(\mathbf{s}_t)$, we shall define a corresponding random element $(\mathbf{s}_\xi)$ in $\mathcal{M}([0,1],S)$ by letting $\mathbf{s}_\xi = \mathbf{s}_{\lfloor \xi T \rfloor}$ for each $\xi \in [0,1]$. In this way, $(\mathbf{s}_\xi)$ "looks like" the original sample path.

The Product Topology

If X is a topological space, then any product X^I is by default endowed with the product topology. This principle is universal throughout this dissertation, regardless of what kind of notation is used, or of any other considerations.

Convergence in the product topology is pointwise convergence, so if $(x_{i,\eta})$ is an asymptotic sequence in X^I, and $(y_i) \in X^i$ is fixed, then

$$(x_{i,\eta}) \to (y_i)$$

means that $x_{i,\eta} \to y_i$ as $\eta \to \infty$ for each fixed coordinate $i \in I$,

The other topologies we will use are the ℓ_1 topology on $\mathbb{R}^\omega$, and the uniform topology on $\mathcal{M}([0,1],X)$, and we will explicitly denote convergence in either as follows.

- $x_\eta \xrightarrow{\ell_1} y$ denotes ℓ_1 convergence in $\mathbb{R}^\omega$;
- $(x_{\xi,\eta}) \xrightarrow{\infty} (y_\xi)$ denotes uniform convergence in $\mathcal{M}([0,1],X)$, for any metric space X.

Although there are many other possible topologies (i.e. the uniform topology on $\mathbb{R}^\omega$), we will work almost exclusively with just these.

2.2 Pointwise and ℓ_1 Convergence

In this section we discuss the relationship between pointwise and ℓ_1 convergence in the countable product space $\mathbb{R}^\omega$. Our convention, as described above, is to denote elements of $\mathbb{R}^\omega$ in "function notation," and the countable coordinate space is $\mathbb{Z}^* = \{0,1,\ldots\}$, so an element of $\mathbb{R}^\omega$ is written as

$$x = (x(0), x(1), \ldots).$$

Of course, the choice of $\mathbb{Z}^*$ as coordinate space is clearly without loss of generality, and we may equivalently use any countable set.

By default, $\mathbb{R}^\omega$ is endowed with the product topology, so $x_\eta \to x$ indicates pointwise convergence; this means that, for each fixed i, $x_\eta(i) \to x(i)$. The other topology we will use is the ℓ_1 topology, which is the topology induced by the ℓ_1 norm:

$$\|x\|_1 = \sum_{i=0}^{\infty} |x| \,.$$

We will denote the normed space of all x such that $\|x\|_1 < \infty$ by the same symbol

$$\ell_1 = \{x \in \mathbb{R}^\omega : \|x\|_1 < \infty\}.$$

Convergence with respect to the ℓ_1 norm is denoted by $x_\eta \xrightarrow{\ell_1} x$, which means that $\|x_\eta - x\|_1 \to 0$. Note that ℓ_1 convergence is well-defined in all of $\mathbb{R}^\omega$, but we are generally only interested in ℓ_1 convergence within the space $\ell_1 \subseteq \mathbb{R}^\omega$ itself.

2.2.1 Summability

Definition 2.2.1. An asymptotic sequence x_η in ℓ_1 is *summable* if:

1. $\limsup_{n\to\infty} \|x_\eta\|_1 < \infty$;
2. $\lim_{J\to\infty} \limsup_{\eta\to\infty} \sum_{i>J} |x_\eta(i)| = 0$.

The idea is that, for a summable sequence, pointwise convergence implies ℓ_1 convergence. Before proving this, we make some simple observations.

Proposition 2.2.1. *Assume x_η is summable. Then:*

1. *the asymptotic sequence y_η defined by $y_\eta(i) = |x_\eta(i)|$ is summable;*
2. *if z_η is summable, then so is $y_\eta = x_\eta + z_\eta$;*
3. *if $x \in \ell_1$ is fixed, and c is any constant, then $y_\eta = x + c \cdot x_\eta$ is summable;*

4. *if y_η satisfies $\sum_{i>J} |y_\eta(i)| < \sum_{i>J} |x_\eta(i)|$ for all η and all J (including $J < 0$), then y_η is summable;*

5. *if y_η satisfies $|y_\eta(i)| < |x_\eta(i)|$ for all η and all i, then y_η is summable;*

6. *all of the above hold if the pre-condition is assumed to hold just for all $\eta > N$, where N is a fixed constant.*

Proof. The first two claims are immediate by definition. For the third claim clearly $c \cdot x_\eta$ is summable, as is any constant asymptotic sequence $(x, x, \ldots)$ in ℓ_1, so the third claim follows from the second.

For the fourth claim, by choosing $J < 0$, we have $\|y_\eta\|_1 \leq \|x_\eta\|)1$, and the second summability condition is satisfied by y_η by definition. The assumption that $|y_\eta(i)| < |x_\eta(i)|$ for all i from the fifth claim implies the assumption $\sum_{i>J} |y_\eta(i)| < \sum_{i>J} |x_\eta(i)|$ from the fourth claim, which in turn implies summability.

Finally, it is clear that the preconditions only need to hold for $\eta > N$, since any finite segment of an asymptotic sequence does not affect limits. □

Proposition 2.2.2. *Let x_η be an asymptotic sequence in ℓ_1, let $x \in \ell_1$ be fixed, and assume $x_\eta \to x$. Then $x_\eta \xrightarrow{\ell_1} x$ if and only if x_η is summable.*

Proof. By considering $y_\eta = x_\eta - x$ if necessary, we may assume w.l.o.g. that the limit is the zero vector $x = 0 = (0, 0, \ldots)$. The forward implication is now trivial, since $x_\eta \xrightarrow{\ell_1} 0$ implies that $\|x_\eta\|_1 \to 0$, which in turn implies that $\limsup_{\eta\to\infty} \sum_{i>J} |x_\eta(i)| = 0$ for all J.

For the reverse implication, pointwise convergence implies that, for any fixed J, $\sum_{i\leq J} |x_\eta(i)| \to \sum_{i\leq J} |x(i)|$, which is equal to 0 due to the assumption that x is the zero vector. We now compute

$$\begin{aligned}\limsup_{\eta\to\infty} \|x_\eta\|_1 &= \lim_{J\to\infty} \Big(\limsup_{\eta\to\infty} \sum_{i\leq J} |x_\eta(i)| + \limsup_{\eta\to\infty} \sum_{i>J} |x_\eta(i)| \Big) \\ &= 0 + \lim_{J\to\infty} \sum_{i>J} |x_\eta(i)|\end{aligned}$$

which is equal to 0 by summability. □

2.2.2 Equivalence of Topologies

We now establish that ℓ_1 convergence is equivalent to pointwise convergence combined with convergence of ℓ_1 norm.

Proposition 2.2.3. *Let x_η be an asymptotic sequence in $\mathbb{R}^\omega$, and let $x \in \mathbb{R}^\omega$ be fixed, and assume $x_\eta \to x$. Then:*

1. $\liminf_{\eta\to\infty} \|x_\eta\|_1 \geq \|x\|_1$,
2. *if $\|x\|_1 < \infty$, then $x_\eta \xrightarrow{\ell_1} x$ if and only if $\|x_\eta\|_1 \to \|x\|_1$.*

Proof. For any fixed J, pointwise convergence implies that $\lim_{\eta\to\infty} \sum_{i\leq j} |x_\eta(i)| = \sum_{i\leq J} |x(i)|$, and therefore we may again compute

$$\begin{aligned}\limsup_{\eta\to\infty} \|x_\eta\|_1 &= \lim_{J\to\infty} \Big(\sum_{i\leq J} |x(i)| + \limsup_{\eta\to\infty} \sum_{i>J} |x_\eta(i)| \Big) \\ &= \|x\|_1 + \lim_{J\to\infty} \limsup_{\eta\to\infty} \sum_{i>J} |x_\eta(i)| \,.\end{aligned}$$

Since $\lim_{J\to\infty} \limsup_{\eta\to\infty} \sum_{i>J} |x_\eta(i)| \geq 0$, the first claim has been proved.

For the second claim, the forward implication is trivial, since ℓ_1 convergence clearly implies $\|x_\eta\|_1 \to \|x\|_1$. For the reverse implication, the above equation shows that if $\|x_\eta\|_1 \to \|x\|_1 < \infty$, then $\lim_{J\to\infty} \limsup_{\eta\to\infty} \sum_{i>J} |x_\eta(i)| = 0$, which means that $x_\eta(i)$ is summable and therefore converges to x in ℓ_1. □

We may restate this proposition in terms of topological equivalence.

Corollary 2.2.4. *The ℓ_1 topology, when restricted to the space $\ell_1 \subseteq \mathbb{R}^\omega$, is equivalent to the topology induced by the product topology and the mapping $x \mapsto \|x\|_1$ from $\ell_1 \to \mathbb{R}$.*

Proof. Since the countable product of metric topologies is metrizable, this is an immediate consequence of proposition 2.2.3. □

2.2.3 The Space $\Phi = \mathrm{Dist}(\mathbb{Z}^*)$

The space of distributions on $\mathbb{Z}^* = \{0, 1, \ldots\}$ carries special significance in this dissertation, since it will serve to parametrize the random graphs we study. We note shall denote this space by

$$\Phi = \mathrm{Dist}(Z^*),$$

and we note that Φ is a subset of ℓ_1, and in fact

$$\Phi = \{\mu \in \ell_1 : \|\mu_1\| = 1, \mu(i) \geq 0 \text{ for all } i\}.$$

The equivalence of pointwise and ℓ_1 convergence in Φ is an immediate corollary to proposition 2.2.3.

Corollary 2.2.5. *If μ_η is an asymptotic sequence in Φ, and $\mu \in \Phi$ is fixed, then $\mu_\eta \to \mu$ if and only if $\mu_\eta \xrightarrow{\ell_1} \mu$.*

Proof. Since every element of Φ has $\|\mu\|_1 = 1$, then clearly $\|\mu_\eta\|_1 \to \|\mu\|_1 = 1$ regardless of pointwise convergence. Hence, by proposition 2.2.3 if $\mu_\eta \to \mu$ then $\mu_\eta \xrightarrow{\ell_1} \mu$ (and the revers implication is trivial). □

We may again state this in terms of topological equivalence.

Corollary 2.2.6. *The product and ℓ_1 topologies coincide on Φ.*

Proof. Immediate. □

2.3 Large Deviations

In this section, we present some preliminary classical results regarding large deviation probabilities. We shall not require any sophisticated results from the theory of large deviations (as in e.g. [25]). Instead, we shall be satisfied with upper bounds of the form

$$\mathsf{P}[L_n] \leq e^{-Cn}$$

for a sequence of events $(L_n)_{n=1}^{\infty}$. Moreover, we are not concerned with optimizing the value of the constant C in the above equation; it will suffice for our purposes to prove that such a constant exists. We are also not interested in lower bounds on large deviation probabilities. As such, the results discussed below are quite elementary, and they are not stated in the strongest possible form.

The following theorem is essentially a restatement of Chernoff's classical upper bound on large deviation probabilities.

Theorem 2.3.1 (Chernoff [19])**.** *For any $\epsilon, \delta > 0$ and any continuous function $f : [0, \epsilon] \to \mathbb{R}$ satisfying $f(0) = 0$ and $f'(0) \leq 0$, there exists a constant $C > 0$ such that the following statement holds.*

Let $(\mathbf{x}_t)$ be an $\mathbb{R}$-valued process with $\mathbf{x}_0 = 0$ and such that

$$\ln \mathsf{E}_t[e^{z\Delta \mathbf{x}_{t+1}}] \leq f(z)$$

always for all t and all $z \in [0, \epsilon]$, where $\Delta \mathbf{x}_{t+1} = \mathbf{x}_{t+1} - \mathbf{x}_t$. Then, for all t, we have

$$\mathsf{P}\left[\mathbf{x}_t \geq \delta t\right] \leq e^{-Ct}.$$

Proof. First, note that for any $z \in [0, \epsilon]$ and any $t > 0$, we have

$$\mathsf{E}\left[e^{z\mathbf{x}_t}\right] = \mathsf{E}\left[e^{z\mathbf{x}_{t-1}} e^{z\Delta \mathbf{x}_t}\right] = \mathsf{E}\left[e^{z\mathbf{x}_{t-1}} \cdot \mathsf{E}_{t-1}\left[e^{z\Delta \mathbf{x}_t}\right]\right] \leq \mathsf{E}\left[e^{z\mathbf{x}_{t-1}}\right] f(z).$$

It thus follows inductively that $\mathsf{E}\left[e^{z\mathbf{x}_t}\right] \leq t \cdot f(z)$.

For any $\delta > 0$, and $z \in [0, \epsilon]$, Markov's inequality now yields

$$\mathsf{P}\left[\mathbf{x}_t \geq \delta t\right] = \mathsf{P}\left[e^{z\mathbf{x}_t} \geq e^{z\delta t}\right] \leq \mathsf{E}\left[e^{z(\mathbf{x}_t - \delta t)}\right] \leq e^{(f(z) - \delta z)t}.$$

Moreover, since $f(0)$ and $f'(0) \leq 0$, then for z sufficiently small, we have $f(z) < \delta z$, and hence for such a z, we choose $0 < C < f(z) - \delta z$, and conclude

$$\mathsf{P}\left[\mathbf{x}_t \geq \delta t\right] \leq e^{-Ct}.$$

□

2.3.1 Independent Random Variables

We now derive two corollaries which deal with sums of i.i.d. random variables. Recall that for any distribution μ on $\mathbb{R}$, $\mathbf{x}_\mu$ denotes a generic μ-distributed random variable, and $\sum_{i=1}^t \mathbf{x}_{\mu,i}$ denotes the sum of t independent copies of $\mathbf{x}_\mu$.

Corollary 2.3.2. *Let μ be a distribution on $\mathbb{R}$ such that $|M(\mu)| < \infty$ and $\mathsf{E}[e^{\epsilon \mathbf{x}_\mu}] < \infty$ for some $\epsilon > 0$. Then for any $M > M(\mu)$, there exists a constant $C > 0$ such that, for all t, we have*

$$\mathsf{P}\left[\sum_{i=1}^t \mathbf{x}_{\mu,t} > Mt\right] \leq e^{-Ct}.$$

Proof. For each t, let $\mathbf{y}_t = \mathbf{x}_{\mu,t} - M(\mu)$, and let $\mathbf{Y}_t = \sum_{i=1}^t \mathbf{y}_t$, noting that $\Delta \mathbf{Y}_t = \mathbf{y}_t$. Next, recall that the moment-generating function can be written as

$$g(z) = \mathsf{E}[e^{z\mathbf{x}_\mu}] = \sum_{i=0}^\infty z^i \mathsf{E}[\mathbf{x}_\mu^i],$$

and therefore $g(0) = 1$ and $g'(0) = \mathsf{E}[\mathbf{x}_\mu] = M(\mu)$. Moreover, by assumption $g(\epsilon) < \infty$, and therefore $g(z)$ is continuous and finite for $z \in [0, \epsilon]$.

Next, we compute the cumulant-generating function of a random variable $\mathbf{y} \stackrel{d}{=} \mathbf{x}_\mu - M(\mu)$ by

$$\begin{aligned} f(z) = \ln \mathsf{E}[e^{z\mathbf{y}}] &= \ln \mathsf{E}[e^{z(\mathbf{x}_\mu - M(\mu))}] \\ &= \ln \mathsf{E}[e^{z\mathbf{x}_\mu}] - zM(\mu) \\ &= \ln g(z) - zM(\mu). \end{aligned}$$

It follows that $f(z) = 0$ and $f'(z) = g'(z)/g(z) - M(\mu) = 0$.

Since

$$\ln \mathsf{E}_t[e^{z\Delta \mathbf{Y}_{t+1}}] = \ln \mathsf{E}_t[e^{z\mathbf{y}_{t+1}}] = f(z)$$

by definition, then by Theorem 2.3.1, C can be chosen such that $\mathsf{P}[\mathbf{Y}_t > (M - M(\mu))t] < e^{-Ct}$ and the proof is complete. □

An easy way to bound the moment-generating function is if the distribution μ is itself bounded.

Corollary 2.3.3. *Let μ be a distribution on $\mathbb{R}$ and assume that $|M(\mu)| < \infty$, for some fixed K, $\mathsf{P}[\mathbf{x}_\mu > K] = 0$. Then for any $M > M(\mu)$, there exists a constant $C > 0$ such that, for all t, we have*

$$\mathsf{P}\left[\sum_{i=1}^{t} \mathbf{x}_{\mu,t} > Mt\right] \leq e^{-Ct}.$$

Proof. The bound $\mathsf{P}[\mathbf{x}_\mu > K] = 0$ implies that $\mathsf{E}[e^{z\mathbf{x}_\mu}] \leq e^{zK} < \infty$ for all $z \geq 0$. □

We make two final comments. First, clearly the signs can be reversed to achieve upper rather than lower bounds. Second the assumption that $|M(\mu)| < \infty$ is not strictly necessary; the bound on the moment-generating function ensures that $M(\mu) < \infty$, and if $M(\mu) = -\infty$, then the above inequalities will in fact hold for any $M \in \mathbb{R}$. We omit this argument since we do not need this result, but the proof is straightforward.

2.3.2 Azuma's Martingale Concentration Inequality

An $\mathbb{R}$-valued process $(\mathbf{x}_t)$ is a *martingale* if $\mathsf{E}_t[\Delta\mathbf{x}_{t+1}] = 0$, and a *supermartingale* if $\mathsf{E}_t[\Delta\mathbf{x}_{t+1}] \leq 0$; note that any supermartingale is also a martingale. Our next corollary is Azuma's inequality for supermartingales with bounded differences.

Corollary 2.3.4 (Azuma's Inequality [6])**.** *For all $\epsilon, C_1 > 0$ there exists $C_2 > 0$ such that, for any supermartingale $(\mathbf{x}_t)$ with $\mathbf{x}_0 = 0$ and such that $|\Delta\mathbf{x}_t| < C_1$ always for all t, we have, for all t,*

$$\mathsf{P}\left[|\mathbf{x}_t| > \epsilon t\right] \leq e^{-C_2 t}.$$

Proof. Without loss of generality, we may assume the bound on increments is $|\mathbf{x}_t| \leq 1$. In this case, by convexity, any random variable $\mathbf{x}$ such that $\mathsf{E}[\mathbf{x}] \leq 0$ and $|\mathbf{x}| \leq 1$ satisfies

$$\mathsf{E}[e^{z\mathbf{x}}] \leq \frac{e^{-z} + e^{z}}{2} = \cosh(z).$$

Hence, such a martingale satisfies $\mathsf{E}_t[e^{z\Delta\mathbf{x}_{t+1}}] \leq \cosh(z)$, and since $\cosh(0) = 1$ and $\cosh'(0) = 0$, the bound follows from Theorem 2.3.1. □

Chapter 3

Random Graphs and the Configuration Model

In this chapter, we present a detailed introduction to the *configuration model*, which will serve as our basic random graph model throughout this dissertation. The configuration model was originally developed by Bender and Canfield [9] and Bollobás [11] as a means of generating a random graph with a prescribed sequence of vertex degrees. This model is not quite as simple as the Erdős-Rényi models $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$, and a certain amount of bookkeeping is required at the outset before we can smoothly manipulate random graphs generated by the configuration model. However, this initial investment in technical bookkeeping ultimately pays dividends, since the configuration model is more versatile and expressive than the Erdős-Rényi models; in fact, several other random graph models, including the Erdős-Rényi models, can be simulated by the configuration model.

Chapter Organization

In section 3.1 we introduce the configuration model, along with some basic related definitions and notation. In section 3.2, we describe the *algorithmic method*, which is our principal analytical technique for studying random graphs generated by the configuration model. In sections 3.3 and 3.4, we discuss the problem of parametrizing the configuration model asymptotically. Finally, in section 3.5, we discuss how the configuration model can be used to simulate other random graph models, and we discuss certain specific instances of the configuration model which carry special significance.

History and Background

The configuration model was originally developed by Bender and Canfield [9] and Bollobás [11], and its earliest applications were to the study of random regular graphs.

The algorithmic method is a common technique in random graph theory, particularly for the configuration model. Perhaps the earliest celebrated algorithmic result is that of Karp and Sipser [43] regarding maximum matchings and independent sets in $\mathbf{G}_{n,p}$. An overview of algorithmic results for random graphs can be found in [35].

Asymptotic parametrization by degree distribution was used by Molloy and Reed [55, 56] in the analysis of the giant component. Similar parametrizations have been used by other authors, in Kim [46] and [2].

3.1 Graphs and Configurations

Recall that a *graph* is a pair (V, E) where V is a set of vertices, and E is a set of edges, each of which is an unordered pair of vertices. In this dissertation, we will primarily work with a similar combinatorial structure called a *configuration*, which is defined as follows.

Definition 3.1.1. A *configuration* is a triple (A, V, E), in which:

- A is a set of *endpoints*;
- V is a set of *vertices*, and each endpoint $a \in A$ belongs to a single vertex $V(a) \in V$;
- E is a set of *edges*, each of which contains two distinct endpoints, and such that E forms a perfect matching of A.

Graphs and configurations are closely related, and given a configuration $G = (A, V, E)$, we may obtain a graph $\hat{G} = (V, \hat{E})$ by creating a graph-edge $\hat{e} = \{V(a_1), V(a_2)\} \in \hat{E}$ for each configuration-edge $e = \{a_1, a_2\} \in E$. A graph can thus be understood as an equivalence class of configurations, and the mapping $G \mapsto \hat{G}$ as a canonical projection. The relationship between graphs and configurations is illustrated in figure 3.1

In order to generate a random configuration, we begin with just the pair (A, V), and choose the edge set uniformly at random:

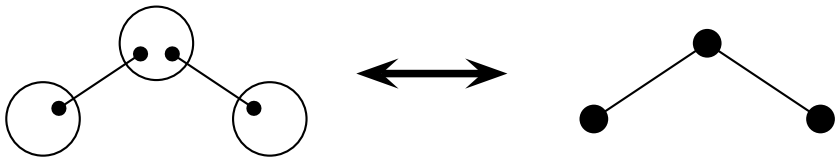


Figure 3.1: On the left, a configuration $G = (A, V, E)$, and on the right, the corresponding graph $\hat{G} = (V, \hat{E})$.

- an *endpoint partition* is a pair $H = (A, V)$ such that each $a \in A$ belongs to a single $V(a) \in V$;[1]
- $\boldsymbol{\mathcal{E}}(A)$ denotes a uniformly random matching of any (even) set A;
- $\boldsymbol{\mathcal{G}}(H) = (A, V, \boldsymbol{\mathcal{E}}(A))$ denotes the corresponding random configuration.

As described above, we may obtain a random graph via canonical projection $\boldsymbol{\mathcal{G}}(H) \mapsto \hat{\boldsymbol{\mathcal{G}}}(H)$, and this method of generating a random graph is called the *configuration model*. We now introduce basic notation, along with certain key definitions, related to the configuration model.

3.1.1 Notation and Conventions

The configuration model will serve as our exclusive random graph model throughout this dissertation, and therefore, at a technical level, we will generally work directly with configurations rather than graphs. However, most "natural" combinatorial properties depend only on the ordinary graph $\hat{G}$, and thus in most cases, we will implicitly identify any configuration G with its image $\hat{G}$ under canonical projection.

Accordingly, in general settings, the term *graph* will be used to refer to either a graph or a configuration, and the term *ordinary graph* will mean specifically a graph and not a configuration. We will use the symbol G for configurations or graphs, and we may refer to

[1] This definition reflects the fact that the endpoint set A is partitioned into subsets $A(v) = \{a \in A : V(a) = v\}$ by the vertex set.

the structure $\boldsymbol{\mathcal{G}}(H)$ as a *random graph*. Also, we may say that an edge *connects* two vertices even though the things being connected are actually endpoints, and so on.

One noteworthy departure from convention is that we shall generally denote the number of endpoints and vertices in a configuration $G = (A, V, E)$ by

$$n = n(G) = |V| \text{ and } \quad m = m(G) = |A|\,,$$

respectively. The number of edges is therefore $|E| = m/2$, which differs from the usual convention whereby the number of edges in graph is denoted by m. We will typically reserve the symbols "n" and "m" for this purpose exclusively.

Basics

These following three notations are essential:

- $A(v)$ denotes the set of endpoints which belong to a vertex $v \in V$;
- $V(a)$ denotes the vertex which contains an endpoint $a \in A$;
- $E(a)$ denotes the edge which contains an endpoint $a \in A$;

Relations Among Endpoints

We will typically adopt an "endpoint-centric" approach to configurations, meaning that structural details will generally be described in terms of relationships between endpoints rather than vertices or edges. Intuitively, for example, we will think of a "location" in a configuration as an endpoint rather than a vertex, and we shall use the following notation to describe the other nearby endpoints:

- $\vec{E}(a) = E(a)\backslash\{a\}$ denotes the endpoint matched to a (and therefore $E(a) = \{a, \vec{E}(a)\}$);
- $\vec{V}(a) = A(V(a))\backslash\{a\}$ denotes the set of other endpoints belonging to the same vertex as a.

Endpoint and Vertex Subsets

The above definitions will often be used in conjunction with the convention that, the image of a subset $S_0 \subseteq S$ under a mapping $f : S \to R$ is denoted by $f(S_0) = \{f(s_0) : s_0 \in S_0\} \subseteq R$. For example, given an endpoint subset $A_0 \subseteq A$, we have the following abbreviations:

- $E(A_0) = \{E(a_0) : a_0 \in A_0\}$ denotes the set of all edges which contain at least one $a_0 \in A_0$;
- $V(A_0) = \{V(a_0) : a_0 \in A_0\}$ denotes the set all vertices which contain at least one $a_0 \in A_0$.

Sub-Structures

We will loosely refer to any combination of subsets of the endpoint, vertex, and edge sets of a configuration (A, V, E) as a *sub-structure*. There are many kinds of sub-structures and we will not attempt to catalog them here; for now, we only give some basic definitions related to endpoint subset $A_0 \subseteq A$:

- (A_0, V) is an endpoint partition, in which all endpoints keep their vertex assignments;
- A_0 is *edge-closed* if $E(A_0)$ is a perfect matching of A_0, or equivalently, if $\vec{E}(A_0) = A_0$;
- if A_0 is edge-closed, then $(A_0, V, E(A_0))$ is a configuration.

The fact that (A_0, V) is (obviously) an endpoint partition reveals a potential ambiguity in the definition, which we briefly discuss. As the name suggests, in an endpoint partition (A, V), the the endpoint set A is partitioned into disjoint subsets $A(v)$ by the vertex set. However, the actual vertex v is not the "same object" as the set $A(v)$ of endpoints it contains. In contrast to an edge in a configuration, which is "just" an unordered pair of endpoints, a vertex has its own identity beyond its endpoint set.

In an endpoint partition (A_0, V) where $A_0 \subseteq A$, each $v \in V$ now contains a (possibly) different set of endpoints:

$$A_0(v) = A_0 \cap A(v) = \{a_0 \in A_0 : V(a) = v\}.$$

However, the actual vertex v retains its original identity in spite of this fact.

At a technical level, then, the use of "containment" terminology to describe the relationship between vertices and endpoints is slightly abusive. In fact, an endpoint partition is not just a pair (A, V), but a triple consisting of an endpoint set A, a vertex set V, and a function mapping each endpoint to its assigned vertex, which we are also denoting with the symbol V. Nevertheless, for intuitive purposes, we shall continue to say that v "contains" an endpoint a if $V(a) = v$.

3.1.2 Degrees and Residual Degrees

The *degree* of a vertex v in an ordinary graph (V, E) is the number of edges incident on v, counting multiplicities. In a configuration (A, V, E), the degree of a vertex v can be computed simply by counting endpoints. As mentioned above, we will typically adopt an "endpoint-centric" view of a configuration, and the appropriate corresponding quantity is the *residual degree*, which is defined as follows.

Definition 3.1.2. For a configuration $G = (A, V, E)$:

- the *degree* of a vertex $v \in V$ is the number of endpoints it contains:

$$\deg(v) = |A(v)|,$$

- the *residual degree* of an endpoint $a \in A$ is one less than the degree of the corresponding vertex:

$$\mathrm{res}(a) = 1 - \deg(V(a)).$$

We will give an intuitive justification for the residual degree below; first, we present two more crucial definitions. We will frequently characterize the multiset $\{\deg(v)\}_{v \in V}$ of

vertex degrees (or the multiset $\{\text{res}(a)\}_{a \in A}$ of residual degrees) in terms of probability distributions as follows.

Definition 3.1.3. For a configuration $G = (A, V, E)$:

- the *degree distribution* λ_G is the distribution of the random variable $\deg(\mathbf{v})$, where $\mathbf{v}$ chosen uniformly at random from V:

$$\lambda_G(i) = \frac{|\{v \in V : \deg(v) = i\}|}{|V|}, \tag{3.1}$$

- the *residual distribution* μ_G is the distribution of the random variable $\text{res}(\mathbf{a})$, where $\mathbf{a}$ is chosen uniformly at random from A:

$$\mu_G(i) = \frac{|\{a \in A : \text{res}(a) = i\}|}{|A|}. \tag{3.2}$$

For a configuration (A, V, E), degrees and residual degrees do not depend on the edge set. Hence, the degree and residual distribution are well-defined with respect to an endpoint partition (A, V) alone, and we shall denote these by $\lambda_{(A,V)}$ and $\mu_{(A,V)}$, respectively.

The Residual Degree

While the degree of a vertex is a fairly natural parameter, the residual degree of an endpoint is perhaps, at first glance, less intuitively appealing. However, in many cases, the residual degree turns out to be a more useful quantity, and we now discuss why this is the case.

As noted above, we will often adopt an "endpoint-centric" view of a configuration. Of course, the notion of degree can naturally be generalized to an endpoint a by just counting the total number of endpoints belonging to the corresponding vertex $V(a)$. This quantity, which we shall call the *true degree*, is indeed useful in certain situations; we shall denote the true degree simply by

$$\deg(a) = \deg(V(a)) = |A(V(a))|\,.$$

The residual degree, which is one less than the true degree, is best understood as the total number of endpoints belonging to $V(a)$ *other than the given endpoint* a. As defined in §3.1.1, this endpoint set is denoted by $\vec{V}(a)$, and therefore the residual degree can be expressed as

$$\mathrm{res}(a) = \left|\vec{V}(a)\right| = \left|A(V(a))\backslash\{a\}\right|.$$

The term "residual" is derived from the fact that, if the given endpoint a were removed, the degree of the vertex $V(a)$ would then be equal to the residual degree $\mathrm{res}(a)$.

The significance of the residual degree can be illustrated as follows. Suppose we are performing an ordinary search on a configuration $G = (A, V, E)$, beginning at a vertex $v_1 \in V$, and in our first step, we traverse an edge $e = \{a_1, a_2\}$ to reach a second vertex $v_2 = V(a_2)$. Now, although the vertex v_2 has a total of $\deg(v_2) = |A(v_2)|$ incident edges, we have already traversed one of these edges in order to reach v_2. Hence the residual degree of the endpoint a_2 is precisely the number of additional edges which we might choose to explore without backtracking.

In other words, the "branching factor" of our search tree is determined by the residual degrees, rather than the true degrees, of the endpoints explored during the search. For instance, if $\mathrm{res}(a_2) = 0$, then we have reached a "dead end," if $\mathrm{res}(a_2) = 1$, then there is only one way to continue this search, and so on. This is illustrated in figure 3.2

For similar reasons, the residual distribution generally provides a better characterization of the local structure of a random configuration $\boldsymbol{\mathcal{G}}(A, V)$ than the degree distribution. In a random configuration, the endpoints are matched uniformly at random, and therefore any fixed $a_1 \in A$ will be matched to a random neighbor $\mathbf{a}_2$, which is uniformly distributed on the set $A\backslash\{a_1\}$. And, if we were to traverse this edge, the (random) number of additional edges we can now explore without backtracking will be (more or less) distributed according to the residual distribution $\mu_{(A,V)}$ of the endpoint partition (A, V).

As a result, the local structure of a $\boldsymbol{\mathcal{G}}(A, V)$ will resemble a recursive random tree structure called a *Galton-Watson tree* [4] generated by the residual distribution. This local

resemblance is one of the definitive characteristics of random configurations, and is discussed more rigorously in chapters 7 and 8.

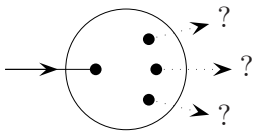


Figure 3.2: The residual degree corresponds to the "branching factor" in a search, which in this case is equal to 3.

The Degree and Residual Distributions

For any configuration G (or any endpoint partition H), the residual distribution can be easily computed from the degree distribution as follows. First, each vertex of degree i contains i endpoints of residual degree $i-1$ (i.e. of true degree i). And, the fraction $|A| / |V|$ is equal to the average vertex degree, which is the first moment of the degree distribution, and which we denote by

$$M(\lambda_G) = \sum_{i=0}^{\infty} i\lambda_G(i).$$

This leads to the formula

$$\mu_G(i) = \frac{(i+1)\lambda_G(i+1)}{M(\lambda_G)}. \tag{3.3}$$

On the other hand the degree distribution λ_G cannot be recovered from μ_G, since the addition of isolated vertices (vertices of degree 0) does not affect the residual distribution. Nevertheless, the relative values of $\lambda_G(i)$ for $i \geq 1$ are uniquely determined by μ_G, and thus if we are given the fraction $\lambda_G(0)$ of isolated vertices, we may compute

$$\lambda_G(i) = \left(\frac{\mu_G(i-1)}{i}\right)\left(\frac{1-\lambda_G(0)}{\sum_{j=1}^{\infty}\frac{\mu_G(j-1)}{j}}\right). \tag{3.4}$$

Moreover, isolated vertices have a trivial effect on the structure of a graph, and hence the residual distribution μ_G effectively characterizes the relevant portion of the degree distribution λ_H.

Unscaled Parameters

The degree distribution is scaled by dividing by the total number of vertices, but in some cases, we will need to work with the raw number of vertices of a given degree. We shall denote this quantity by

$$\Lambda_G(i) = |\{v \in V : \deg(v) = i\}|,$$

and we will refer to the set of values $\Lambda_G(i)$ for $i \geq 0$ as the *unscaled degree distribution*; note also that, clearly, we have $\lambda_G(i) = \frac{\Lambda_G(i)}{|V|}$.

Another useful unscaled parameter is the *degree sequence*. In some cases, we will assume that the vertex set V is canonically ordered in some way, in which case we shall denote the sequence of vertex degrees, according to this ordering, by

$$\mathcal{D}_G = (\deg(v_1), \deg(v_2), \ldots, \deg(v_n)).$$

Since the order of vertices has a negligible effect on the structure of a graph, the degree sequence rarely provides any useful information which is not available from the degree distribution. But, for technical reasons, it is sometimes necessary to deal with such issues as canonical orderings.

3.1.3 Universal Sets

For both high-level discussion and technical details, we will often work with universal sets:

- $\mathbb{G}$ denotes the set of all (finite) configurations;
- $\mathbb{H}$ denotes the set of all (finite) endpoint partitions;

A reader concerned that these "sets" might not exist is referred to section 2.1.4; the idea is not to get into set-theory, but rather to be able to describe a "property" of configurations as a mapping $\alpha : \mathbb{G} \to \{\top, \bot\}$, and so on.

We shall also use blackboard letters to relate such sets to each other as follows:

- $\mathbb{H}(G) = (A, V)$ denotes the endpoint partition of a given $G = (A, V, E) \in \mathbb{G}$;
- $\mathbb{G}(H) = \{G \in \mathbb{G} : \mathbb{H}(G) = H\}$ denotes the set configurations with endpoint partition H;
- $\mathbb{E}(A)$ denotes the set of all perfect matchings of a finite set A.[2]

To illustrate how these definitions are useful, for any Boolean property of the form $\alpha : \mathbb{G} \to \{\top, \bot\}$, we may induce a mapping $P_\alpha : \mathbb{H} \to [0, 1]$ by

$$P_\alpha(H) = \mathsf{P}[\alpha(\boldsymbol{\mathcal{G}}(H)) = \top] = \frac{|\{G \in \mathbb{G}(H) : \alpha(G) = \top\}|}{|\{\mathbb{G}(H)\}|}.$$

The behavior of the property α with respect to the configuration model can be succinctly described in terms of this function.

Distributions on $\mathbb{G}$

Using the terminology from §2.1.1, the symbol $\boldsymbol{\mathcal{G}}$ can be understood as a *random mapping* from $\mathbb{H}$ to $\mathbb{G}$. The image $\boldsymbol{\mathcal{G}}(H)$ of this random mapping is thus a *generic* random element in $\mathbb{G}$, which is an equivalence class modulo equivalence in distribution.

In some cases, we will encounter a random configuration $\mathbf{G}$ which is not directly generated in this way. However, in this case, we may still analyze $\mathbf{G}$ using similar techniques, provided that the distribution of $\mathbf{G}$, conditional on its endpoint partition $\mathbb{H}(\mathbf{G})$ (which is a random element in $\mathbb{H}$), is the uniform distribution on $\mathbb{G}(\mathbb{H}(\mathbf{G}))$. This condition can be stated as follows.

Definition 3.1.4. A random configuration $\mathbf{G}$ is $\mathbb{H}$-*conditionally uniform* if, for every $H \in \mathbb{H}$ such that $\mathsf{P}[\mathbb{H}(\mathbf{G}) = H] > 0$, and every $G \in \mathbb{G}(H)$, we have

$$\mathsf{P}[\mathbf{G} = G \mid \mathbb{H}(\mathbf{G}) = H] = \frac{1}{|\mathbb{G}(H)|}.$$

[2]Note that $\mathbb{E}(A) = \varnothing$ unless $|A|$ is even, since an odd set cannot be perfectly matched; similarly, $\mathbb{G}(H) = \varnothing$ unless the number of endpoints is even, but in general, this technicality can be overlooked.

In "probabilistic" notation, this condition can be expressed as

$$\mathbf{G} \stackrel{d}{=} \boldsymbol{\mathcal{G}}(\mathbb{H}(\mathbf{G})).$$

The right-hand side of the above expression denotes a random element in $\mathbb{G}$ which is constructed by first taking the random endpoint partition $\mathbb{H}(\mathbf{G})$, and then applying the random mapping $\boldsymbol{\mathcal{G}}$, which returns a uniformly random configuration with this endpoint partition.

We may also express this condition explicitly in terms of the distribution of $\mathbf{G}$. We denote the uniform distribution on $\mathbb{G}(H)$ by $\mathfrak{U}(\mathbb{G}(H))$, which is the distribution of $\boldsymbol{\mathcal{G}}(H)$:

$$\mathfrak{D}[\boldsymbol{\mathcal{G}}(H)] = \mathfrak{U}(\mathbb{G}(H)).$$

Hence, an $\mathbb{H}$-conditionally uniform distribution is a mixture of such uniform distributions:

$$\mathfrak{D}[\mathbf{G}] = \sum_{H \in \mathbb{H}} \mathsf{P}[\mathbb{H}(\mathbf{G}) = H] \cdot \mathfrak{U}(\mathbb{G}(H)).$$

3.1.4 Symmetry and Enumeration of $\mathbb{E}(A)$

The "source" of randomness in the configuration model is the uniformly random matching $\mathbf{E}$ of the endpoint set A, and in this section, we examine some some basic properties of the set $\mathbb{E}(A)$ of perfect matchings of a given set A.

Unlike the Erdős-Rényi random graph model $\mathbf{G}_{n,p}$, the occurrences of edges in a uniformly random matching are not independent events. As a result, certain common probabilistic techniques involving sums of random variables are not as easily applicable to the configuration model. On the other hand, the set $\mathbb{E}(A)$ exhibits various symmetries which are not generally available in independent-edge models.

In particular, the uniform distribution on $\mathbb{E}(A)$ is clearly invariant under permutations of the endpoint set. This invariance has several immediate implications, the simplest of which is the following.

Proposition 3.1.1. *Let* $\mathbf{E}$ *be a uniformly random perfect matching of* A*, and choose any* $a \in A$*. Then:*

1. $\vec{\mathbf{E}}(a)$ *is a uniformly random element in* $A\backslash\{a\}$*;*

2. *conditional on* $\vec{\mathbf{E}}(a)$*, the remaining edge set* $\mathbf{E}\backslash\mathbf{E}(a)$ *is a uniformly random matching of the set* $A\backslash\{a, \vec{\mathbf{E}}(a)\}$*.*

Proof. Immediate by symmetry. □

More generally (and somewhat trivially), the distribution of $\mathbf{E}$ conditional on any given property being satisfied is uniform on the set of perfect matchings which satisfy this property. These symmetries can be established without explicitly dealing with probabilities; they are simply consequences of the fact that any given matching is just as likely to occur as any other. Ultimately, of course, actual probabilities will necessarily enter into the picture. However, as a general guideline (or perhaps an aesthetic principle), it is preferable to exploit the symmetry of the uniform distribution on $\mathbb{E}(A)$ rather than to explicitly compute probabilities.

Needless to say, this guideline does not hold universally. Indeed, although the edge occurrences in a uniformly random matching are not strictly independent, these dependencies (for disjoint endpoint pairs) are very slight. And, in most cases, this "near independence" can be substituted for true independence to achieve similar results. Nevertheless, this sort of technique runs somewhat against the grain, and, even in cases where these independence-based techniques seem appealing, it is often possible to find a simpler solution by exploiting symmetries.

Enumeration of Perfect Matchings

The number of perfect matchings of an m-element set can be computed easily using standard combinatorial techniques. We may identify each perfect matching of the set $[m] = \{1, \ldots, m\}$ with an equivalence class of permutations by defining, for any permutation $\sigma : [m] \to [m]$, the corresponding matching

$$\mathcal{E}(\sigma) = \{\{\sigma_1, \sigma_2\}, \{\sigma_3, \sigma_4\}, \ldots, \{\sigma_{m-1}, \sigma_m\}\}$$

Accordingly, any permutation acts on the set of perfect matchings of $[m]$ by composition, so $\rho(\mathcal{E}(\sigma)) = \mathcal{E}(\rho \circ \sigma)$, and thus for any matching E, we have

$$\rho(E) = \{\{\rho_i, \rho_j\} : \{i, j\} \in E\}.$$

The total number of perfect matchings can be computed by counting symmetries as follows.

Proposition 3.1.2. *For any even m, the number of perfect matchings of an m-element set is*

$$(m-1)!! = (m-1)(m-3)\cdots 1.$$

Proof. The group of symmetries is of any given matching is generated by transpositions of two elements in the same edge, along with permutations of the entire edge set. This group of symmetries thus contains $2^{m/2}(m/2)!$ elements, and thus the total number of perfect matchings is

$$\begin{aligned}\frac{m!}{(m/2)!2^{m/2}} = \frac{m(m-1)\cdots 1}{(m/2)(m/2-1)\cdots 1 \cdot 2^{m/2}} &= \frac{m(m-1)\cdots 1}{2\cdot(m/2)\cdot 2\cdot(m/2-1)\cdots 2\cdot 1} \\ &= \frac{m(m-1)\cdots 1}{m(m-2)\cdots 2} \\ &= (m-1)(m-3)\cdots 1.\end{aligned}$$

□

The quantity $(m-1)!! = (m-1)(m-3)\ldots 1$ is the *double factorial* of $(m-1)$. We note that

$$m! = (m-1)!! \cdot m!!,$$

and hence the double factorial can be roughly approximated by the square root of the ordinary factorial.

The double factorial formula has a simple interpretation which will play a key role in our algorithmic methods. Specifically, we may choose a canonical representative permutation for each perfect matching as follows. First, we require that the first element in the permutation is $\sigma_1 = 1$, and thus we have $(m-1)$ choices for the second element σ_2. Then,

we require that the σ_3 is the smallest of the remaining elements so $\sigma_3 = 2$ unless $\sigma_2 = 2$, in which case $\sigma_3 = 3$. We then have $(m-3)$ choices for σ_4, and so on, so total number of choices is $(m-1)(m-3)\cdots 1 = (m-1)!!$.

3.1.5 Configurations and Simple Graphs

Although configurations are arguably interesting combinatorial structures in their own right, ordinary graphs are of much more traditional and perhaps "natural" structures. Of course, given a configuration $G = (A, V, E)$, we may obtain an ordinary graph on the same vertex set V via the *projection mapping* $G \mapsto \hat{G} = (V, \hat{E})$. The set of graph-edges $\hat{E}$ is the multiset

$$\hat{E} = \{V(a_1), V(a_2)\}_{e=\{a_1,a_2\}\in E}$$

of unordered pairs of vertices.

Accordingly, the configuration model may be used to generate an ordinary graph by $\hat{\mathbf{G}} = \hat{\boldsymbol{\mathcal{G}}}(A, V)$, and the degree of each vertex in this graph will be the same as its degree $\deg(v) = |A(v)|$ in the endpoint partition (A, V). However, there are certain potential issues which warrant discussion.

First, by definition $\boldsymbol{\mathcal{G}}(A, V)$ is uniformly distributed on the set $\mathbb{G}(A, V)$ of all configurations with endpoint partition (A, V). However, this does not immediately imply anything about the distribution of the corresponding random ordinary graph.

Second, recall an ordinary graph (V, E) is *simple* if

1. each edge connects two distinct vertices (i.e. there are no *loops*);
2. no two edges connect the same pair of vertices (i.e. there are no *parallel edges*).

We are typically interested in simple graphs, but the configuration model allows for the possibility of loops and parallel edges.

For the purpose of discussing the relationship between configurations and simple graphs, we will temporarily work with a canonically ordered vertex set $V = (v_1, \ldots, v_n)\}$,

in which case an endpoint partition $H = (A, V)$ can be specified canonically by the degree sequence $\mathcal{D}_H = (\deg(v_1), \ldots, \deg(v_n))$.

Proposition 3.1.3. *Let $H = (A, V)$ be an endpoint partition with degree sequence $\mathcal{D}_H = (d_1, \ldots, d_n)$, and assume that $\sum_{i=1}^{n} d_i = m$ is even. Then there are exactly $\prod_{i=1} d_i!$ configurations in $\mathbb{G}(A, V)$ which correspond to each simple graph with degree sequence $\mathcal{D}$.*

Proof. We proceed by induction on twice the number of endpoints (i.e. the number of edges). For $m = 2$, this is trivial, since there is only one configuration and one such simple graph. Now inductively assume the claim holds for $m - 2$, choose any simple graph $\hat{G} = (V, \hat{E})$, and pick any edge $\hat{e}$.

W.l.o.g, we may assume that $\hat{e} = \{v_1, v_2\}$; note that since $\hat{G}$ is simple then this edge cannot be a loop. Also, clearly, there are $d_1 \cdot d_2$ ways to choose endpoints a_1, a_2 from each of v_1, v_2 to form a corresponding configuration edge $\{a_1, a_2\}$. Since $\hat{G}$ is simple, then $\hat{e}$ is the unique graph-edge connecting v_1 and v_2, and therefore exactly one such pair $\{a_1, a_2\}$ of endpoints must belong to every perfect matching which generates $\hat{G}$.

Finally, by induction, for each pair $\{a_1, a_2\}$ there are $(d_1 - 1)!(d_2 - 1)! \prod_{i=3}^{n} d_i!$ perfect matchings $A \backslash \{a_1, a_2\}$ which yield the remaining set of graph edges $\hat{E} \backslash \{\hat{e}\}$, and therefore the total number of matchings which yield $\hat{E}$ is $\prod_{i=1}^{n} d_i!$. □

Corollary 3.1.4. *Every simple graph with degree sequence $\mathcal{D}_H$ occurs with equal probability as $\hat{\boldsymbol{\mathcal{G}}}(H)$.*

Proof. Immediate. □

Due to this fact, the configuration model can be used to generate uniformly random simple graphs, provided that the probability of generating a simple graph is not too low. Computing this probability in general is rather difficult, but for sparse graphs, by imposing maximum degree constraints or bounds on the moments of the degree distribution, the probability of encountering loops or multiple edges can be kept under control.

We shall not get into details, because our primary concern is the random configuration $\boldsymbol{\mathcal{G}}(A, V)$ itself. In the appendix we demonstrate that the configuration model can be used to simulate the Erdős-Rényi random graph $\mathbf{G}_{n,m}$ by conditioning on simplicity in this way. For a more detailed discussion on such topics as enumeration of simple graphs, we refer the reader to, e.g. [52, 11, 9].

3.2 The Algorithmic Method

The *algorithmic method* refers loosely to a collection of techniques for studying random combinatorial structures algorithmically. While there are many ways to accomplish this, we may roughly classify algorithmic techniques as either *constructive* or *destructive*.

The *constructive* method involves generating a random graph $\boldsymbol{\mathcal{G}}(A, V)$ algorithmically. In this case, the initial state is the endpoint partition (A, V), and as time passes, edges are added incrementally until ultimately the algorithm terminates with a complete, uniformly random graph $\boldsymbol{\mathcal{G}}(A, V)$. Due to the symmetries discussed above, there are many algorithmic techniques for generating a uniformly random matching. Therefore, for a given graph property α, it is often possible to devise a constructive algorithm such that this property can be determined fairly easily from the execution path.

The *destructive* method involves predicting the output of an ordinary graph algorithm on a random input graph. That is, given a graph property α and an algorithm which computes α, we would attempt to analyze the execution path of this algorithm when the initial state is the random graph $\boldsymbol{\mathcal{G}}(A, V)$. The term "destructive" derives from the fact that the algorithms we will analyze typically modify the structure of graph by, for example, removing edges. In this case, the execution path will be somewhat of a "mirror image" of a constructive algorithm, in that edges are incrementally removed over time, and the terminal state is just the endpoint partition (A, V).

As we shall see, the constructive and destructive methods are closely related, and in most cases, we can express a given constructive algorithm destructively and vice versa. In either case, an essential characteristic of the algorithmic method is that the "randomness"

of a random graph $\boldsymbol{\mathcal{G}}(A, V)$ is uncovered in small increments. The algorithmic method thus stands in contrast to what we may call "static" methods, whereby one attempts to reason about the entire structure of a random graph at once.

The configuration model is particularly suited for algorithmic analysis due to the symmetries of the set $\mathbb{E}(A)$ of perfect matchings of the endpoint set A. Indeed, while "static" methods are often based on concentration inequalities and sums of independent random variables, the success of the algorithmic method hinges on the ability to successfully manage the incremental release of information. For this purpose, the symmetry and the uniform randomness of the configuration model are often more valuable than independence.

3.2.1 Algorithmic Generation of a Random Graph

We first introduce the constructive algorithmic method, which involves generating a random configuration $\boldsymbol{\mathcal{G}}(A, V)$ algorithmically. The configuration model is based on the uniformly matching $\mathbf{E}$ of the endpoint set A, which can be generated by a simple recursive algorithm as follows.

Proposition 3.2.1. *The following recursive algorithm generates a uniformly random matching of an (even) set A:*

1. *choose a (possibly random) endpoint* $\mathbf{a}_1 \in A$ *arbitrarily;*
2. *choose a second endpoint* $\mathbf{a}_2 \in A \backslash \mathbf{a}_1$ *uniformly at random;*
3. *create an edge* $\mathbf{e} = \{\mathbf{a}_1, \mathbf{a}_2\}$*, let* $\mathbf{A} = A \backslash \{\mathbf{a}_1, \mathbf{a}_2\}$*, and return the matching*

$$\mathbf{E} = \{\mathbf{e}\} \cup \boldsymbol{\mathcal{E}}(\mathbf{A}).$$

Proof. Immediate by symmetry. □

We shall call this basic algorithm the *configuration model (CM) algorithm*. Typically we will analyze the CM algorithm in iterative, rather than recursive form. In this case, we begin with an endpoint set $\mathbf{A}_0 = 0$, and at each time step, we remove a single endpoint

$\mathbf{a}_t \in \mathbf{A}_{t-1}$ to yield $\mathbf{A}_t = \mathbf{A}_{t-1} \backslash \{\mathbf{a}_t\}$. The method of choosing the endpoint thus depends on whether t is even or odd:

- at each *odd* time t, we may choose $\mathbf{a}_t$ arbitrarily from $\mathbf{A}_{t-1}$;
- at each *even* time t, $\mathbf{a}_t$ must be chosen uniformly at random from $\mathbf{A}_{t-1}$;
- after each *even* time step, we create an edge $\mathbf{e}_t = \{\mathbf{a}_{t-1}, \mathbf{a}_t\}$.

Note that this algorithm can be customized by specifying the method by which odd numbered endpoints are selected. In applications, the odd numbered endpoints will be selected in a manner which facilitates the analysis of a particular graph property.

3.2.2 Destructive Methods

The destructive algorithmic method involves predicting the output of a particular graph algorithm on a random input graph. Since we are only concerned with the output, we will adopt a somewhat primitive definition of an "algorithm" which suits our needs.

Definition 3.2.1. A *graph algorithm* (or, more precisely, a *configuration algorithm*) consists of:

1. a *transition function* $\pi : \mathbb{G} \to \mathbb{G}$;
2. a set of *terminal states* $\mathbb{G}_T \subseteq \mathbb{G}$.

In addition:

- the *execution path* for any input graph G is the sequence $(G_0, G_1, \ldots, G_T)$, where $G_0 = G$, G_T is a terminal state (assuming such a state is reached), and $G_{t+1} = \pi(G_t)$ for each t;
- this graph algorithm *computes* a property $\alpha : \mathbb{G} \to \{\top, \bot\}$ if $\alpha(G) = \alpha(\pi(G))$ for all $G \in \mathbb{G}$.

We note that there are various "loopholes" in this definition. For instance, it may be the case that a terminal state is never reached. Also, an "algorithm" of this sort need not bear any resemblance to an actual computational procedure. The actual algorithms we encounter, though, will not suffer from any of these problems, and since the above definition meets our requirements, we shall not dwell on potential technicalities.

At an intuitive level, what is important is that an algorithm transforms a given configuration G in such a way that the property $\alpha : \mathbb{G} \to \{\top, \bot\}$ is preserved. Ideally, $\pi(G)$ should be "simpler" than the original configuration G, and thus as the algorithm executes, the problem of determining $\alpha(G) = \alpha(G_t)$ becomes progressively easier, until a terminal state is reached, at which point the property $\alpha(G_T)$ becomes "trivial."

Random Input and the Observable Process

If the input graph is random, then the execution path $(\boldsymbol{\mathcal{G}}_0, \boldsymbol{\mathcal{G}}_1 = \pi(\boldsymbol{\mathcal{G}}_0), \ldots, \boldsymbol{\mathcal{G}}_{\boldsymbol{\tau}})$ will also be random, as will the termination time $\boldsymbol{\tau}$. However, this sequence is a "random process" in name only, since the entire execution path is predictable from the initial state. The "randomness" of the initial random graph $\boldsymbol{\mathcal{G}}_0$ is thus released all at once, which more or less defeats the purpose of constructing such a random process in the first place.

In order to achieve an incremental release of randomness, we shall restrict our access to information as follows:

- the *observable state* at any time t is the endpoint partition $\mathbf{H}_t = \mathbb{H}(\mathbf{G}_t)$;
- the $\mathbb{H}$-valued process $(\mathbf{H}_t) = (\mathbb{H}(\mathbf{G}_t))$ is called the *observable process*.

Unlike the actual execution path, the observable process will indeed be "morally" random. The observable process somewhat corresponds to the experience of monitoring the execution of an actual algorithm on a real computer. In this case, the true state is the internal state of the computer: the exact value stored in every individual memory register, etc. This is, of course, more information than a human can process, so instead one must

choose certain key parameters to monitor, which hopefully provide an overall picture of the internal memory state.

Uniformity Preserving Algorithms

Without imposing restrictions on the transition function π, this observable process may not be "well-behaved." Recall that, as defined in §3.1.3, the distribution of random configuration $\mathbf{G}$ is $\mathbb{H}$-*conditionally uniform* if, conditional on its endpoint partition $\mathbf{H} = \mathbb{H}(\mathbf{G})$, the distribution of $\mathbf{G}$ is uniform on the set $\mathbb{G}(\mathbf{H})$. We this offer the following definition.

Definition 3.2.2. A graph algorithm is *uniformity preserving* if:

1. for any endpoint partition H, the distribution of $\pi(\boldsymbol{\mathcal{G}}(H))$ is $\mathbb{H}$-conditionally uniform;
2. there exists a subset $\mathbb{H}_T \subseteq \mathbb{H}$ such that any state G is terminal if and only if $\mathbb{H}(G) \subseteq \mathbb{H}_T$.

It is fairly evident that these conditions ensure that the observable process will be Markov. We shall describe the corresponding transition kernel in "probabilistic" notation by defining the *observable transition function* to be the random mapping $\boldsymbol{\pi} : \mathbb{H} \to \mathbb{H}$ defined by

$$\boldsymbol{\pi}(H) \stackrel{d}{=} \mathbb{H}(\pi(\boldsymbol{\mathcal{G}}(H))). \tag{3.5}$$

Intuitively, this means that, for any $H \in \mathbb{H}$, $\boldsymbol{\pi}(H)$ is a random element in $\mathbb{H}$, which is distributed identically to the endpoint partition $\mathbb{H}(\pi(\boldsymbol{\mathcal{G}}(H)))$ of the random graph $\pi(\boldsymbol{\mathcal{G}}(H))$. By convention, each call to random function is assumed to be independent, and can thus be used to define a Markov chain, in the natural way.

Proposition 3.2.2. *For a uniformity preserving graph algorithm, and any $H \in \mathbb{H}$, the observable process with initial state $\boldsymbol{\mathcal{G}}(H)$ is a stopped, $\mathbb{H}$-valued Markov chain for which:*

1. *the transition probabilities satisfy $\mathbf{H}_{t+1} = \boldsymbol{\pi}(\mathbf{H}_t)$, where $\boldsymbol{\pi}$ is defined in* (3.5)*;*

2. the stopping time is $\boldsymbol{\tau} = \min\{t : \mathbf{H}_t \in \mathbb{H}_T\}$.

Proof. We shall first inductively show that

$$\mathfrak{D}[\mathbf{G}_t \mid \mathbf{H}_0, \ldots, \mathbf{H}_t] = \mathfrak{D}[\mathbf{G}_t \mid \mathbf{H}_t] = \mathfrak{D}[\boldsymbol{\mathcal{G}}(\mathbf{H}_t)] \tag{3.6}$$

holds for all t (with probability 1). The base case is immediate by assumption, so assume that (3.6) holds for t, and we may extend to $t+1$ as follows.

By definition, (3.6) implies that $\mathfrak{D}[\mathbf{G}_{t+1} \mid \mathbf{H}_0, \ldots, \mathbf{H}_t] = \mathfrak{D}[\pi(\boldsymbol{\mathcal{G}}(\mathbf{H}_t)]$, and by preservation of uniformity,

$$\pi(\boldsymbol{\mathcal{G}}(\mathbf{H}_t)) \stackrel{d}{=} \boldsymbol{\mathcal{G}}(\mathbb{H}(\pi(\boldsymbol{\mathcal{G}}(\mathbf{H}_t)))) \stackrel{d}{=} \boldsymbol{\mathcal{G}}(\boldsymbol{\pi}(\mathbf{H}_t)).$$

It follows that

$$\mathfrak{D}[\mathbf{G}_{t+1} \mid \mathbf{H}_0, \ldots, \mathbf{H}_t] = \mathfrak{D}[\mathbf{G}_{t+1} \mid \mathbf{H}_t] = \mathfrak{D}[\boldsymbol{\mathcal{G}}(\boldsymbol{\pi}(\mathbf{H}_t))],$$

which implies both that (3.6) holds for $t+1$, and also that

$$\mathfrak{D}[\mathbf{H}_{t+1} \mid \mathbf{H}_0, \ldots, \mathbf{H}_t] = \mathfrak{D}[\mathbf{H}_{t+1} \mid \mathbf{H}_t] = \mathfrak{D}[\boldsymbol{\pi}(\mathbf{H}_t)],$$

which proves the first claim.

The second claim is immediate, since by definition 3.2.2, the termination condition can be detected from the observable state by the condition $\mathbf{H}_{\boldsymbol{\tau}} = \mathbb{H}(\mathbf{G}_\tau) \in \mathbb{H}_T$. □

Due to this proposition, once we have established that a given algorithm preserves uniformity, it is no longer necessary to keep track of the actual execution path, since the true state at any time t is simply the uniform distribution on $\mathbb{G}(\mathbf{H}_t)$. Moreover, the fact that the graph algorithm π computes a property α can also be characterized in terms of this observable transition function. As in §3.1.3, we define a function $P_\alpha : \mathbb{H} \to [0,1]$ by

$$P_\alpha(H) = \mathsf{P}[\alpha(\boldsymbol{\mathcal{G}}(H)) = \top].$$

The uniformity preservation property in this case yields the relationship

$$P_\alpha(H) = \mathsf{E}[P_\alpha(\boldsymbol{\pi}(H))], \tag{3.7}$$

and, hence, by tracing the observable process until it terminates, we have

$$P_\alpha(H) = \mathsf{E}[P_\alpha(\mathbf{H}_1)] = \ldots = \mathsf{E}[P_\alpha(\mathbf{H}_{\boldsymbol{\tau}})].$$

And, the termination condition will typically be such that the property α can be determined trivially, so either $P_\alpha(\mathbf{H}_{\boldsymbol{\tau}})$ is equal to either 0 or 1 (for instance, a common termination condition is that $\mathbf{H}_{\boldsymbol{\tau}}$ is empty).

Intuitively, then, the observable process can itself be considered to be the execution path of a (randomized) algorithm for which the state space is the set $\mathbb{H}$ of endpoint partitions. This algorithm does not compute a "property" in the same way as the underlying graph algorithm. However, we are able to compute the value $P_\alpha(H)$ in expectation, in the sense that $\mathsf{E}[P_\alpha(\mathbf{H}_{\boldsymbol{\tau}})] = P_\alpha(H)$.

Simple Edge Removal and Endpoint Choice

The basic destructive method outlined above is limited to algorithms which preserve uniformity in distribution, conditional on the endpoint partition. Intuitively, a uniformity preserving algorithm will make certain local structural modifications to a configuration $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{G}}(H)$, while leaving the remainder of the configuration "undisturbed." Typically, these algorithms will be quite simple, and involve some kind of local search or greedy optimization. However, with a little creativity, it is often possible to express some fairly effective algorithmic techniques in a manner which preserves uniformity.

The easiest way to preserve uniformity is to remove one edge at a time. Removing an edge $e = \{a_1, a_2\}$ from a configuration $G = (A, V, E)$, also requires removing both endpoints a_1, a_2 from the endpoint set A, and we will abbreviate this procedure by writing

$$G - e = (A \backslash \{a_1, a_2\}, V, E \backslash \{e\}).$$

Now, we cannot remove "just any" edge from G if we wish to preserve conditional uniformity. In a sense, we must choose the edge e based on information available in the endpoint partition (A, V). For this purpose, we shall make use of an *endpoint choice function*

which is simply a mapping which chooses an endpoint $a_{\mathcal{C}}(H) \in A$ given an endpoint partition $H = (A, V)$. As discussed in §2.1.4, this has nothing to do with the axiom of choice; the idea is simply to ensure that the choice of endpoint does not depend on the edge set of a configuration.

Proposition 3.2.3. *For any endpoint choice function $a_{\mathcal{C}}$ on $\mathbb{H}$, the following preserves uniformity:*

1. *for any $G \in \mathbb{G}$, choose an endpoint $a_1 = a_{\mathcal{C}}(\mathbb{H}(G))$;*
2. *set $\pi(G) = G - E(a_1)$, where $E(a_1)$ is the edge incident on a_1.*

The observable transition function is given by

$$\boldsymbol{\pi} : (A, V) \mapsto (A\backslash\{a_1, \mathbf{a}_2\}, V),$$

where $a_1 = a_{\mathcal{C}}(A, V)$, and $\mathbf{a}_2$ is a uniformly random element in $A\backslash\{a_1\}$.

Proof. Let $\mathbf{G} = \boldsymbol{\mathcal{G}}(H)$, and note that the choice of a_1 depends only on $H = (A, V)$, and hence a_1 is not random. The edge $\mathbf{e}$ is of course random, as is the endpoint matched to a_1, which we may denote by $\mathbf{a}_2 = \vec{\mathbf{E}}(a_1)$. And, by proposition 3.1.1, $\mathbf{a}_2$ is uniformly distributed in $A\backslash\{a_1\}$, and the restricted matching $\mathbf{E}\backslash\mathbf{E}(a)$ is a uniformly random matching of $A\backslash\{a_1, \mathbf{a}_2\}$ by symmetry. Hence, $\pi(\mathbf{G})$ is a uniformly random configuration on $\mathbf{H}_1 = (A\backslash\{a_1, \mathbf{a}_2\}, V)$, and therefore conditional uniformity is preserved. □

Note that this algorithm is more or less identical to the basic constructive CM algorithm described in §3.2.1. If we keep track of time in terms of individual endpoint removals, rather then edge removals, we once again have a situation in which, at odd time steps, we remove an endpoint deterministically (according to the choice function), while at even time steps, we remove an endpoint uniformly at random.

The constructive and destructive methods are thus closely related, and the difference between the two is perhaps philosophical, and reflects two different conceptions of what it

means for something to be "random." In the constructive case, the edge set does not "exist" until the algorithm executes, while in the destructive case, the entire edge set exists at the outset, but we are not "aware" of it as observers. At a formal level, though, there is no real difference between these two situations, and we may choose whichever interpretation we find more appealing.

Although the constructive method was easier to "set up," in many cases, the destructive method is ultimately more intuitive. This is because it is more natural to describe graph properties in terms of an actual graph algorithm, rather than an algorithm which constructs a configuration. Using the destructive method, we can thus establish the correctness of an algorithm in the usual way, and once we have established that such an algorithm preserves uniformity, we can "let it run," and move on to the analysis of the observable process.

3.2.3 The Removal Process

Both the constructive and destructive algorithms described above are such that the vertex set remains unchanged throughout the execution. In either the case, the key random structure is the following:

- a *removal process* is a random descending chain $(\mathbf{A}_t)$ of subsets of the endpoint set A;
- a removal process is *simple* if $|\mathbf{A}_t| - |\mathbf{A}_{t+1}| = 1$ for all t;
- in this case, the process can be described by the individual endpoint removed each step:

$$\mathbf{A}_{t+1} = \mathbf{A}_t \backslash \{\mathbf{a}_{t+1}\}.$$

Our algorithmic analysis of the configuration model is largely based on endpoint removal processes, along with the corresponding random sequence of $\mathbf{H}_t = (\mathbf{A}_t, V)$. And, in most cases, by decomposing iterations of an algorithm appropriately into individual endpoint removal steps, we shall be able to work with a simple process, which is advantageous for several reasons.

And, as described above in §3.1.2 an endpoint partition can be characterized by its degree distribution (or residual distribution), along with the number of vertices. At a technical level, then, we will primarily work with either $(\boldsymbol{\lambda}_t) = (\lambda_{\mathbf{H}_t})$ or $(\boldsymbol{\mu}_t) = (\mu_{\mathbf{H}_t})$, each of which are random processes for which the state space is the set of distributions on $\mathbb{Z}^*$.

Variations

There are many possible variations on this central theme, and we will very briefly mention two possibilities. First, the technique described above is not the only way to preserve uniformity (or, conversely, to generate a uniform random matching). In chapters 10 and 11, we shall use an slightly different method; nevertheless, the observable process will still be an endpoint removal process as described above, just that way endpoints are selected for removal is not the same.

Second, in a literal sense, the destructive method outlined above is only applicable to Boolean graph properties. Of course, if we are given some other kind of property to compute, there are many ways to modify the technique appropriately. We might, for example, augment the state space by including some auxiliary information along with the graph G_t. Or, the relationship between $\alpha(G_t)$ and $\alpha(G_{t+1})$ might not be more complicated than just equality. However, it is not hard to see that any of these variations are relatively easy to accommodate within the same general framework.

3.3 Asymptotic Parametrization

The theory of random graphs is by nature asymptotic, in that we are generally only concerned with random graphs which are "arbitrarily large." For this purpose, we shall construct *asymptotic sequences* $\mathbf{G}_\eta$ of random graphs, and reason about the limit as $\eta \to \infty$. Our asymptotic conventions were discussed in the previous chapter, but briefly, an asymptotic sequence exists solely for the purpose of taking limits, which for random structures means limits in probability. The symbol η serves no purpose other than to index such sequences, and in applications we will generally leave off the indexing subscript, and take the

limit $\eta \to \infty$ implicitly.

Our asymptotic parametrization is based on convergence of the residual distribution (defined in §3.1.2). Hence, the parameter space is the set

$$\Phi = \mathrm{Dist}(\mathbb{Z}^*)$$

of distributions on $\mathbb{Z}^* = \{0, 1, \ldots\}$. By default, we work with the product topology on Φ, so $\mu_\eta \to \mu$ means pointwise convergence $\mu_\eta(i) \to \mu(i)$ for each fixed i, but, since pointwise and ℓ_1 convergence are equivalent in this situation (see section 2.2 from the previous chapter), the following definition could be stated in terms of either.

Definition 3.3.1. For an asymptotic sequence $H_\eta = (A_\eta, V_\eta)$, the expression

$$H_\eta = \mathcal{H}(\mu) \tag{3.8}$$

specifies that both $m_\eta = |A_\eta| \to \infty$ and $\mu_{H_\eta} \to \mu$ as $\eta \to \infty$.

We shall similarly denote a corresponding asymptotic sequence of random graphs by

$$\mathbf{G}_\eta = \boldsymbol{\mathcal{G}}(\mu),$$

so in this case $\mathbf{G}_\eta = \boldsymbol{\mathcal{G}}(H_\eta)$ for $H_\eta = \mathcal{H}(\mu)$.[3] These notations should be understood in the same way as asymptotic "big O" notation, in that the symbol "=" technically denotes set membership rather than equality. In general discussion, we will use the expression $\boldsymbol{\mathcal{G}}(\mu)$ in statements such as

$$\boldsymbol{\mathcal{G}}(\mu) \text{ satisfies } \alpha \text{ with probability } 1 - o(1)$$

to indicate that, for any asymptotic sequence of endpoint partitions H_η satisfying these assumptions, we have $\lim_{\eta\to\infty} \mathsf{P}[\alpha(\boldsymbol{\mathcal{G}}(H_\eta)) = \top] \to 1$

[3]Note that the number of endpoints $|A_\eta|$ is even for this to be well-defined, and we thus maintain this unstated assumption as well.

Other Possible Approaches

The configuration model has been parametrized asymptotically in a number of more or less equivalent ways (perhaps the "standard" would be the Molloy and Reed [55, 56]). Our definitions differ cosmetically from the common asymptotic parametrizations in two ways.

First, we do not assume that the number of vertices in H_η is exactly η, so as not to confuse the index of an asymptotic sequence with a meaningful property of H_η. The assumption $m_\eta \to \infty$ ensures that the graph becomes "large," and it is not difficult to see that the number $n_\eta = |V_\eta|$ must also tend to ∞; in fact, as will be shown below, the number of non-isolated vertices (i.e. vertices of strictly positive degree) will be of the same order of magnitude $\Theta(m_\eta)$.

Second, for reasons which will become clear, the residual distribution μ_{H_η}, rather than the degree distribution λ_{H_η}, is our basic asymptotic parameter. In §3.3.1, we will show that the same effect can be accomplished with the degree distribution modulo small technicalities. Regardless of which route is chosen, the basic reality is that to reason about the configuration model asymptotically, convergence of the residual distribution is a necessary, and, in most cases, also a sufficient assumption.

In addition to these cosmetic differences, in some situations, certain additional assumptions are required, the most common of which are:

- convergence of one or more moments of the residual distribution;
- a bound on the maximum degree.

These additional assumptions are discussed in §**??**

3.3.1 The Degree and Residual Distributions

We are using the residual distribution μ_H to describe the structure of an endpoint partition, and the number of endpoints $m = |A|$ to describe its size. An alternate, and

perhaps more "natural" choice is to use the degree distribution λ_H and the number of vertices $n = |V|$ for these respective purposes.

The only actual difference between these two choices is minor: the residual distribution does not take into account isolated vertices. For technical reasons it is necessary to allow isolated vertices to exist, but they have no substantive effect on the structure of a graph.

Aside from isolated vertices the degree and residual distributions provide the same information. As shown in §3.1.2, the residual distribution can be computed from the degree distribution; we may thus define the residual distribution algebraically as a function of any $\lambda \in \Phi$ by

$$\mu_\lambda(i) = \frac{(i+1)\lambda(i+1)}{M(\lambda)}, \tag{3.9}$$

where $M(\lambda)$ is the first moment; in order for this to be well defined, of course, we must have $0 < M(\lambda) < \infty$.

Due to this relationship between λ and μ_λ, we may restate our convergence assumptions in terms of the degree distribution as follows.

Proposition 3.3.1. *For any $\lambda \in \Phi$ which satisfies $0 < M(\lambda) < \infty$, and any asymptotic sequence H_η of endpoint partitions, the following two statements are equivalent:*

1. *$H_\eta = \mathcal{H}(\mu)$ and $\lambda_{H_\eta}(0) \to \lambda(0)$;*
2. *$n_\eta = |V_\eta| \to \infty$, $\lambda_{H_\eta} \to \lambda$, and $M(\lambda_{H_\eta}) \to M(\lambda)$.*

In this case, we also have $m_\eta = \Theta(n_\eta)$.

Proof. It is immediate from (3.9) that the second condition implies pointwise convergence $\mu_{H_\eta} \to \mu$. Also, note that the average degree is $M(\lambda_{H_\eta}) = m_\eta/n_\eta$, so the fact that $m_\eta = \Theta(n_\eta)$ is a consequence of the convergence of $M(\lambda_{H_\eta}) \to M(\lambda)$; this in turn implies that $m_\eta \to \infty$, and therefore $H_\eta = \mathcal{H}(\mu)$.

For the opposite implication, note that, given the value $\lambda(0)$, the rest of the degree distribution can be recovered from μ_λ from (3.9) (as shown in (3.4) from §3.1.2) by

$$\lambda(i) = \left(\frac{\mu_\lambda(i-1)}{i}\right)\left(\frac{1-\lambda(0)}{\sum_{j=1}^{\infty}\frac{\mu_\lambda(j-1)}{j}}\right). \tag{3.10}$$

Since $\lambda_{H_\eta}(0) \to \lambda(0)$ and $\mu_{H_\eta}(i-1) \to \mu_\lambda(i-1)$ by assumption, it suffices to show that

$$\sum_{j=1}^{\infty}\frac{\mu_{H_\eta}(j-1)}{j} \to \sum_{j=1}^{\infty}\frac{\mu_\lambda(j-1)}{j}.$$

Since $\sum_{j=0}^{\infty}\mu_{H_\eta}(j) \to \sum_{j=0}^{\infty}\mu(j) = 1$, this follows from dominated convergence (see §2.2.1). □

This proposition is evidently not applicable unless we are given the value $\lim_{\eta\to\infty}\lambda_{H_\eta}(0) = \lambda(0)$. However, for equally obvious reasons, we may assume without any real loss of generality that $\lambda(0) = 0$, in which case either both methods of asymptotic parametrization are precisely equivalent. Note also that in this case we have $m_\eta = \Theta(n_\eta)$, which means that the graphs we are dealing are sparse (i.e. the average degree is $\Theta(1)$).

Factorial Moments

The relationship between λ and μ_λ can also be described in terms of moments. For distributions on $\mathbb{Z}^*$, it is generally preferable to work with factorial moments rather than ordinary moments, and hence, for any integer $k \geq 1$, we shall denote the k'th factorial moment of a distribution λ by

$$M_k(\lambda) = \sum_{i\in\mathbb{Z}^*}(i)_k\lambda(i),$$

where $(i)_k$ denotes the falling factorial $(i)_k = i(i-1)\cdots(i-k+1)$; since we will use the first moment far more often than the higher moments, we will continue to abbreviate $M(\lambda) = M_1(\lambda) = \sum_i i\lambda(i)$.

Proposition 3.3.2. *For any $\lambda \in \Phi$ with $0 < M(\lambda) < \infty$, we have*

$$M_k(\mu_\lambda) = \frac{M_{k+1}(\lambda)}{M(\lambda)}. \tag{3.11}$$

Proof. We may compute

$$M_k(\mu_\lambda) = \sum_{i \in \mathbb{Z}^*} (i)_k \mu_\lambda(i) = \sum_{i \in \mathbb{Z}^*} \frac{(i+1)(i)_k \lambda(i+1)}{M(\lambda)} = \sum_{i \in \mathbb{Z}^*} \frac{(i+1)_k \lambda(i+1)}{M(\lambda)}$$
$$= \frac{M_{k+1}(\lambda)}{M(\lambda)}.$$

□

3.3.2 Other Considerations

We now briefly touch upon certain variations of our standard asymptotic assumptions which will arise on occasion. In addition to technicalities related to isolated vertices discussed above, there are three more substantial issues worthy of mention.

1. We will occasionally require convergence of the first moment $M(\mu_{H_\eta}) \to M(\mu)$. Due to proposition 3.3.2, this can also be accomplished using the degree distribution by $M_2(\lambda_{H_\eta}) \to M_2(\lambda)$.

2. A bound on the maximum degree is sometimes required, the purpose of which is to prevent one vertex, or a small number of vertices, from adversely affecting the structure of the entire graph. For technical reasons we will not get into at this point, the bound we shall use in this situation will be of the form $m_\eta^{1/8-\Omega(1)}$.

3. In some cases, the endpoint partition will itself be random. This occurs, for instance, if we are able to show that $\mathbf{G}_\eta = \boldsymbol{\mathcal{G}}(\mu)$ will contain a sub-graph $\mathbf{G}'_\eta$ which is uniformly distributed, conditional on its endpoint partition. In this case, if we can predict the (random) residual distribution $\mu_{\mathbf{G}'_\eta}$, then we can reason about the structure of $\mathbf{G}'_\eta$ based on results about a uniformly random configuration with this residual distribution. Generally speaking, definition 3.3.1 will handle random parametrization more or less effortlessly, but there are certain technicalities involved, and therefore we postpone this discussion until the next section.

3.4 Asymptotics and Probability

Now that we have defined the asymptotic random graph $\boldsymbol{\mathcal{G}}(\mu)$, the next step is to decide what to "do with it." A natural starting point is to consider a given graph property α, and attempt to compute the limit

$$\lim_{\eta\to\infty} \mathsf{P}[\mathbf{G}_\eta \text{ satisfies } \alpha].$$

In many cases, we find that the above limit is either 0 or 1; this phenomenon is called a *zero-one law*.

Of course, certain graph properties, such as the diameter or the number of connected components, cannot naturally be expressed as a Boolean statement. Instead, these graph properties may take the form of a function β from the set $\mathbb{G}$ of all configurations to $\mathbb{R}$. In this case, we cannot generally determine the exact value of $\beta(\mathbf{G}_\eta)$, and the corresponding phenomenon is convergence in probability, which we shall call *concentration*. Broadly defined, then, our "objective" is to uncover zero-one laws and concentrations, and we now proceed to give more precise definitions regarding both concepts.

3.4.1 High-Probability Guarantees

In many cases, we will be able to determine not only that $\mathbf{G}_\eta$ satisfies a given graph property with probability tending to 1, but also the rate at which this convergence occurs. In general, it is desirable to achieve the strongest possible guarantee on the rate of convergence in probability, and for this purpose, we will use the following high probability guarantees.

Definition 3.4.1. Let L_η be an asymptotic sequence of events, let x_η be an asymptotic sequence in $\mathbb{R}$, and assume $x_\eta \to \infty$. Then:

- L_η occurs *asymptotically almost surely* (a.a.s.) if $\mathsf{P}[L_\eta] = 1 - o(1)$;
- L_η occurs with *x_η-polynomially high probability* (x_η-w.p.h.p.) if $\mathsf{P}[L_\eta] = 1 - x_\eta^{-\Omega(1)}$;
- L_η occurs with *x_η-exponentially high probability* (x_η-w.e.h.p.) if $\mathsf{P}[L_\eta] = 1 - e^{-\Omega(x_\eta)}$;

- L_η occurs with *ultimately always* (u.a.) if there exists a fixed $N \in \mathbb{N}$ such that $\mathsf{P}[L_\eta] = 1$ for all $\eta \geq N$;
- unqualified uses of "w.e.h.p." are assumed to mean m_η-w.e.h.p., where $m_\eta = |A_\eta|$ is the number of endpoints in H_η, and similarly for w.p.h.p.

The phrase "$\mathsf{P}[L_\eta]$ *occurs with high probability*" (w.h.p.) will be used as a placeholder for which any of the above phrases may be substituted. Obviously, all instances of "w.h.p." within a given context must be substituted in the same way.

The two intermediate guarantees, which specify a rate of convergence, depend on some other asymptotic sequence x_η with respect to which this rate is determined. Again, this shields the asymptotic index η from any substantive involvement. Note also that none of these guarantees distinguish between rates of convergence $\mathsf{P}[L_\eta] \to 1$ which differ by a constant factor. This has two important consequences:

- if $n_\eta = \Theta(m_\eta)$ then the conditions m_η-w.h.p. and n_η-w.h.p. are equivalent;
- if both L_η and K_η occur w.h.p., then so does $L_\eta \wedge K_\eta$.

The first of these consequences allows us to interchange the number of vertices or endpoints in H_η for the purposes of defining high-probability, assuming of course, that issues related to isolated vertices do not get in the way. In most cases, though, m_η is more natural, since the duration of random processes associated with the algorithmic method is typically equal to the number of endpoints.

The fact that conjunctions of w.h.p. events occur w.h.p. allows us to perform a sort of logical inference among w.h.p. events. In the next chapter, we develop this idea more fully into what we shall call "w.h.p. logic." For now, our objective is simply to present basic definitions, and thus we will not go into details.

Of course, only *finite* conjunctions of w.h.p. events occur w.h.p. Infinite conjunctions do not generally hold w.h.p., nor do conjunctions where the number of events increases with η. For this reason, we will generally deal with one asymptotic event at a time, and

for a set $\{L_{i,\eta} : i \in I\}$ of asymptotic events, an assertion of the form "$L_{i,\eta}$ occurs w.h.p. for each $i \in I$" means that each individual asymptotic event $L_{i,\eta}$ occurs w.h.p., and not the conjunction of these events. We may violate this convention as follows:

- these events occur *simultaneously w.h.p.* if $K_\eta = \bigwedge_{i \in I} L_{i,\eta}$ occurs w.h.p.;
- these events occur *uniformly w.h.p.* if $P_\eta = \inf\{\mathsf{P}[L_{i,\eta}] : i \in I\}$ tends to 1 at the specified rate.

3.4.2 Convergence in Probability

The asymptotic behavior of non-Boolean random objects cannot be adequately described by the w.h.p. occurrence of any single event, and we must instead consider convergence in probability. In the interest of generality, we will present our basic definitions in the setting of a topological space X, rather than in metric (i.e. ϵ, δ) terms.

Definition 3.4.2. Let $\mathbf{x}_\eta$ be an asymptotic random element in X, and let $x \in X$ be fixed. Then:

- $\mathbf{x}_\eta \to x$ w.h.p. specifies that, for every fixed open $B \ni x$, the event $\mathbf{x}_\eta \in B$ occurs w.h.p.
- $\mathbf{x}_\eta \nrightarrow x$ w.h.p. specifies that there exists a fixed open $B \ni x$ such that $\mathbf{x}_\eta \notin B$ occurs w.h.p.

The condition $\mathbf{x}_\eta \to x$ w.h.p. is called *w.h.p. convergence*, and the condition $\mathbf{x}_\eta \nrightarrow x$ w.h.p. is called *w.h.p. separation*.

The symbol "$\nrightarrow$" in a separation should be understood as meaning "is bounded away from" rather than "does not converge to." Also, it is worth mentioning that the w.h.p. convergences and separations which are satisfied by a given $\mathbf{x}_\eta$ are completely described in terms of w.h.p. occurrences of the asymptotic events $\mathbf{x}_\eta \in B$ and $\mathbf{x}_\eta \notin B$ for all open sets $B \subseteq X$. Hence, although convergence is not a Boolean phenomenon, establishing a w.h.p.

convergence is a matter of showing that, for each individual neighborhood $B \ni x$, the event $\mathbf{x}_\eta \in B$ occurs w.h.p.

Convergence in Metric Spaces

If $\mathbf{x}_\eta$ is sequence of random elements in a metric space (X, ρ), then these definition can be equivalently defined in terms of the metric ρ. For example, we have $\mathbf{x}_\eta \to x$ w.h.p. if, for all $\epsilon > 0$, the event $\rho(\mathbf{x}_\eta, x) < \epsilon$ occurs w.h.p. We may also use these definitions in conjunction with "big O" notation; for instance, given an asymptotic sequence of non-negative random variables $\mathbf{x}_\eta$:

- $\mathbf{x}_\eta = o(1)$ w.h.p. indicates that $\mathbf{x}_\eta \to 0$ w.h.p.;
- $\mathbf{x}_\eta = O(1)$ w.h.p. indicates that $\mathbf{x}_\eta \nrightarrow \infty$;

This notation can be used in the usual algebraic way; for instance, the meaning of $\mathbf{x}_\eta = o(\mathbf{y}_\eta)$ is clear, and so on.

We emphasize that, if a convergence $\mathbf{x}_\eta \to x$ holds w.p.h.p. or w.e.h.p., this places a restriction on the rate of convergence, but this only refers to the rate at which the value $\mathsf{P}[\mathbf{x}_\eta \in B]$ tends to 1 for each *fixed* neighborhood B of x; it has nothing to do with the rate at $\mathbf{x}_\eta$ tends towards x.

For example, consider a sequence $\mathbf{x}_\eta$ of random variables such that that

$$\mathsf{P}[\mathbf{x}_\eta > \epsilon] = e^{-\epsilon m_\eta}$$

for all η and all $\epsilon > 0$. In this situation, for any (non-random) sequence $y_\eta \to 0$, we have

$$\mathsf{P}[\mathbf{x}_\eta < y_\eta] = e^{-m_\eta y_\eta} = e^{-o(m_\eta)},$$

and hence $\mathbf{x}_\eta < y_\eta$ does *not* hold m_η-w.e.h.p. However, for any *fixed* $\epsilon > 0$, we in fact have $\mathsf{P}[\mathbf{x}_\eta < \epsilon]$ w.e.h.p., and therefore the convergence $\mathbf{x}_\eta \to 0$ holds w.e.h.p.

3.4.3 Asymptotic Parametrization Revisited

While an asymptotic random graph $\boldsymbol{\mathcal{G}}(\mu)$ is technically an equivalence class of asymptotic sequences, the construction of such sequences is somewhat of an artifice. The objective is to study random graphs which are both arbitrarily large, and for which the residual distribution is arbitrarily close to μ. Although sequences are perhaps the most expeditious route, the same can be accomplished without this explicit construction.[4]

In order to avoid technicalities, let us define a *normal* graph property to be a mapping $\alpha : \mathbb{G} \to \{\top, \bot\}$ which is not affected by the number of isolated vertices.

Proposition 3.4.1. *For any normal graph property and any $\mu \in \Phi$, the following are equivalent:*

1. *$\boldsymbol{\mathcal{G}}(\mu)$ satisfies α w.e.h.p.;*

2. *there exist $\epsilon, C, M > 0$ such that*

$$\mathsf{P}[\alpha(\boldsymbol{\mathcal{G}}(H)) = \bot] \leq e^{-Cm(H)}. \tag{3.12}$$

holds for any $H = (A, V) \in \mathbb{H}$ satisfying both $m(H) = |A| > M$ and $\|\mu_H - \mu\|_1 < \epsilon$.

Proof. Let us define a function f from $\mathbb{H}$ to the closed interval $[0, \infty]$ by

$$f(H) = -\frac{\ln \mathsf{P}[\alpha(\boldsymbol{\mathcal{G}}(H)) = \bot]}{m(H)},$$

so, by definition, $\boldsymbol{\mathcal{G}}(\mu)$ satisfies α w.e.h.p. if and only if, every asymptotic sequence $H_\eta = \mathcal{H}(\mu)$ satisfies $\liminf_{\eta\to\infty} f(H_\eta) > 0$.

For any $\epsilon > 0$, let us now denote the ϵ-neighborhood of μ with respect to the ℓ_1 norm by B_ϵ, so the second condition asserts that

$$\exists_{\epsilon,C,M} \forall_{H\in\mathbb{H}} \Big((m(H) > M \wedge \mu_H \in B_\epsilon) \implies f(H) \geq C \Big)$$

[4] What makes this possible is the fact that the parameter space Φ is metric. The "correct" way to go about asymptotic parametrization might be to use nets rather than sequences, but in metric spaces this is not necessary.

Since the product and ℓ_1 topology are equivalent on Φ, then any $H_\eta \in \mathcal{H}(\mu)$ must satisfy $\mu_{H_\eta} \xrightarrow{\ell_1} \mu$, along with $m_\eta \to \infty$, and it is immediate that if this condition holds, we must have $\liminf_{\eta\to\infty} f(H_\eta) \geq C > 0$. Hence, the second condition implies the first.

The other implication can be deduced using standard techniques for manipulating sequences in metric spaces. Specifically, the negation of the above logical formula is given by

$$\forall_{\epsilon,C,M}\exists_{H\in\mathbb{H}}\Big(m(H) > M \wedge \mu_H \in B_\epsilon \wedge f(H) < C\Big)$$

Using this fact, we may easily construct a sequence H_η such that $m_\eta \to \infty$ and also $\mu_{H_\eta} \xrightarrow{\ell_1} \mu$ which also satisfies $f(H_\eta) \to 0$. Therefore, if the second of the above conditions fails, it is not the case that $\boldsymbol{\mathcal{G}}(\mu)$ satisfies α w.e.h.p. □

Random Parametrization

The above proposition allows us to accommodate random graphs for which the endpoint partition is itself random. That is, if we are given an asymptotic sequence of random endpoint partitions $\mathbf{H}_\eta$, such that the assumptions in definition 3.3.1 hold w.h.p., the behavior of $\mathbf{G}(\mathbf{H}_\eta)$ will be more or less identical to the case where the parameters are non-random. In this case, the number of endpoints may itself be random, and thus cannot be used to define a high-probability guarantee, but the same effect can be achieved as follows.

Proposition 3.4.2. *Let $\mathbf{H}_\eta$ be an asymptotic sequence of random endpoint partitions, let $z_\eta \to \infty$, and let $\mu \in \Phi$, and assume that*

1. *$\mu_{\mathbf{H}_\eta} \to \mu$ holds z_η-w.h.p.,*
2. *$m(\mathbf{H}_\eta) = \Theta(z_\eta)$ holds z_η-w.h.p.*

Then, the asymptotic random graph $\mathbf{G}_\eta = \boldsymbol{\mathcal{G}}(\mathbf{H}_\eta)$ z_η-w.h.p. satisfies any normal property α which is satisfied w.h.p. by $\boldsymbol{\mathcal{G}}(\mu)$.

Proof. For any such property α, define a function $f : \mathbb{H} \to [0, 1]$

$$f(H) = \mathsf{P}[\alpha(\boldsymbol{\mathcal{G}}(H)) = \perp].$$

Now, since $\mathbf{G}_\eta$ is uniformly distributed conditional on $\mathbf{H}_\eta$, it follows that

$$\mathsf{P}[\alpha(\mathbf{G}_\eta) = \perp] = \mathsf{E}[f(\mathbf{H}_\eta)].$$

For the w.e.h.p. condition, as shown above, the fact that $\boldsymbol{\mathcal{G}}(\mu)$ satisfies α w.e.h.p. implies that there exist constants $M, \epsilon, C > 0$ such that

$$P_\alpha(H) < e^{-Cm(H)}$$

whenever $m(H) > M$ and $\|\mu_H - \mu\|_1 < \epsilon$. Moreover, the two assumptions on $\mathbf{H}_\eta$ guarantee that both $\left\|\mu_{\mathbf{H}_\eta} - \mu\right\|_1 < \epsilon$ and $m(\mathbf{H}_\eta) > C_0 z_\eta$ hold z_η-w.e.h.p., for a fixed constant C_0.

It follows that cases where either of these assumptions fail contribute a total of $e^{-\Omega(z_\eta)}$ to the total of $\mathsf{E}[f(\mathbf{H}_\eta)]$. And, cases where both assumptions hold contribute a maximum of $e^{-CC_0 z_\eta}$ to this expectation.

The argument for the w.p.h.p. condition is more or less identical, and for u.a. the argument is trivial.

For the a.a.s. condition, we must show that $\mathsf{E}[f(\mathbf{H}_\eta)] \to 0$. In this case, the fact that $\boldsymbol{\mathcal{G}}(\mu)$ satisfies α a.a.s. implies that, for any $\epsilon > 0$, there must exist $M, \delta > 0$ such that $f(H) < \epsilon$ whenever both $\|\mu_H - \mu\|_1 < \delta$ and $m(H) > M$. Since $\mathbf{H}_\eta$ must satisfy both of these conditions a.a.s. by assumption, then $\mathsf{E}[f(\mathbf{H}_\eta)] \leq \epsilon + o(1)$, and ϵ can be chosen arbitrarily small. □

3.4.4 Sharp Thresholds

Proposition 3.4.1 also yields insight into the behavior of graph properties with respect to the topology of the parameter space.

Corollary 3.4.3. *For any normal Boolean graph property α, the set*

$$\Phi_\alpha = \{\mu : \boldsymbol{\mathcal{G}}(\mu) \text{ satisfies } \alpha \text{ w.e.h.p.}\}$$

is an open subset of Φ.

Proof. For any $\mu \in \Phi_\alpha$, there exists $\epsilon > 0$ such that the second of the conditions in proposition 3.4.1 is satisfied, and it is evident that $\boldsymbol{\mathcal{G}}(\mu')$ must also satisfy α w.e.h.p. for any μ' in the ϵ-neighborhood of μ. □

In particular, since Φ is connected and thus cannot be expressed as the union of disjoint open sets, then unless either $\Phi_\alpha = \Phi$ or $\Phi_{\neg\alpha} = \Phi$, we cannot hope to determine whether or not $\boldsymbol{\mathcal{G}}(\mu)$ satisfies α w.e.h.p. for every single $\mu \in \Phi$. The best we can do is to find a *sharp threshold*:

- the *threshold* is the set $\Psi_\alpha = \Phi \backslash \{\Phi_\alpha \cup \Phi_{\neg\alpha}\}$;
- this threshold is *sharp* if every every $\mu \in \Psi_\alpha$ belongs to the closure of both Φ_α and $\Phi_{\neg\alpha}$.

Based on our methods of asymptotic parametrization, a sharp threshold is the best possible outcome. Hence once we have found a sharp threshold, we will consider a given problem to have been "solved." This is not to say that threshold behavior is unimportant; on the contrary, the granular details of what structural changes take place while a threshold is "crossed" are often non-trivial and quite interesting. However, this is a different genre of problem, which requires a different methodology and we shall restrict our to determining the location of sharp thresholds.

Concentration

Sharp thresholds are of course only applicable for Boolean graph properties. For $\mathbb{R}$-valued graph properties (or, more generally, graph properties in an arbitrary topological space), the situation is similar, if slightly more complicated. We will not attempt to list all possible asymptotic behaviors in this case, and instead give a brief overview.

First, for a mapping $\beta : \mathbb{G} \to \mathbb{R}$

- Φ_β denotes the set of all $\mu \in \Phi$ such that $\beta(\boldsymbol{\mathcal{G}}(\mu))$ is w.e.h.p. concentrated;
- $F_\beta : \Phi_\beta \to \mathbb{R}$ denotes this w.e.h.p. limit, so $\beta(\boldsymbol{\mathcal{G}}(\mu)) \to F_\beta(\mu)$;

- $\Psi_\mu = \Phi \backslash \Phi_\beta$.

Now, unlike for Boolean graph properties, it is not necessarily the case that Φ_β is open. However, it is not difficult to show that this function F_β must be continuous from $\Phi_\beta \to \mathbb{R}$. Hence, a sharp threshold point in this case corresponds to a distribution $\mu \in \Psi_\beta$ for which F_β is discontinuous (i.e. such that F_β cannot be extend continuously to μ), since in this case it is impossible for $\beta(\boldsymbol{\mathcal{G}}(\mu))$ to converge w.e.h.p. to any value.

It follows that the following two conditions constitute the equivalent of a "sharp threshold," for a continuous graph property

1. the closure of Φ_β is the entire set $\hat{\Phi}$;
2. every $\lambda \in \Psi_\beta$ is a discontinuity point for the function β.

In this case, β cannot be extended continuously to any strict superset of Φ_β, and hence the situation cannot possibly be improved based on our methods for asymptotic parametrization.

Polynomially Versus Exponentially High Probability

The discussion of sharp thresholds above was based on the w.e.h.p. high-probability guarantee, since this is the condition we will use most often. However, it is not difficult to show that, by making the natural adjustments, the same results will hold for the weaker w.p.h.p. guarantee. The same is *not* true for the a.a.s. condition; this is because the rate of converge for both w.e.h.p. and w.p.h.p. involve a fixed constant which we used in proposition 3.4.1. For the a.a.s. condition, the rate of convergence $\mathsf{P}[L_\eta] \to 1$ does not involve a fixed constant.

This is not particularly troubling, since the graph properties are still relatively "well-behaved" with respect to the a.a.s. condition. Moreover, for several reasons, we will work with the two intermediate guarantees w.p.h.p. and w.e.h.p., rather than u.a. and a.a.s. Indeed, the condition u.a., which requires that $\mathsf{P}[L_\eta] = 1$ for n sufficiently large is not

really a "high-probability guarantee," and is only useful to deal with structural and logical restrictions (e.g. the event [if $\mathbf{G}_\eta$ is 4-colorable then $\mathbf{G}_\eta$ is 3-colorable] occurs u.a.).

The weaker condition a.a.s. is used frequently in random graph theory, under the premise that if the probability that $\mathbf{G}_\eta$ satisfies a given property with probability tending to 1, then "almost all" such graphs satisfy this property. In most cases, though, if $\mathbf{G}_\eta$ satisfies a given property a.a.s. the rate of convergence in probability will be either polynomial or exponential. And, the strength of the probability guarantee can usually be inferred from the nature of the problem.

Exponentially high probability generally holds for properties which cannot be affected by changing a small number of edges. For instance, a questions about what fraction of the vertex set of $\mathbf{G}(\mu)$ satisfy a given property will typically be answered by w.e.h.p. convergences. On the other hand, properties which are sensitive to small modifications (i.e. being k-colorable) can usually only be established w.p.h.p.

3.5 Simulating Other Random Graph Models

Our asymptotic parameter space for the configuration model is the space Φ of distributions on $\mathbb{Z}^*$, and most of the techniques in this dissertation will be applicable to all distributions in Φ. Hence, in a broad sense, the objective of this line of research is to uncover the relationship between the residual distribution μ and the structural properties of the random graph $\boldsymbol{\mathcal{G}}(\mu)$.

However, for both practical and mathematical reasons, some particular degree distributions carry special relevance. This is because the configuration model can be used to simulate other random graph models, and we now discuss certain specific distributions which are useful for this purpose.

3.5.1 The Poisson Distribution

We denote a Poisson distribution with expected value c by

$$\pi_c(i) = \frac{e^{-c}c^i}{i!}. \tag{3.13}$$

It is not difficult to show that the degree distribution of the Erdős-Rényi random graph $\mathbf{G}_{n,m}$ with average degree $c = m/2n$ will have Poisson distribution, and using this information, it is possible to derive results about $\mathbf{G}_{n,m}$ (and $\mathbf{G}_{n,p}$) using the configuration model.

The Poisson distribution has the property that the residual distribution is the same Poisson distribution $\mu_{\pi_c} = \pi_c$. This can be easily verified directly from (3.13), or else by factorial moments:

$$\begin{aligned} M_k(\pi_c) = \sum_{i \geq k} (i)_k \pi_c(i) = \sum_{i \geq k} \frac{(i)_k e^{-c}c^i}{i!} &= c^k e^{-c} \sum_{i \geq k} \frac{c^{i-k}}{(i-k)!} \\ &= c^k. \end{aligned}$$

It follows that the factorial moments of the residual distribution μ_{π_c} are also given by

$$M_k(\mu_{\pi_c}) = \frac{M_{k+1}(\pi_c)}{M(\pi_c)} = c^k = M_k(\pi_c),$$

and hence we have $\mu_{\pi_c} = \pi_c$.

It follows that the random graph $\boldsymbol{\mathcal{G}}(\pi_c)$ is more or less the same as the Erdős-Rényi random graph $\mathbf{G}_{n,m}$ with $m/2n = c$, which is in turn more or less the same as $\mathbf{G}_{n,p}$ with $p = c/n$. We shall postpone our rigorous discussion of the relationship between Erdős-Rényi graphs and $\boldsymbol{\mathcal{G}}(\pi_c)$ until we have developed an adequate set of tools for algorithmic analysis. For these reasons, the proof of the following theorem appears in the appendix.

Theorem 3.5.1. *For fixed $c > 0$, let $\boldsymbol{\mathcal{G}}(H) = \boldsymbol{\mathcal{G}}(\pi_c)$ be an asymptotic random configuration with Poisson residual distribution*

$$\pi_c(i) = \frac{e^{-c}c^i}{i!},$$

and also assume that the fraction $\lambda_H(0)$ of vertices of degree 0 converges to $\pi_c(0) = e^{-c}$.

1. *Any property which is satisfied w.e.h.p. by $\hat{\boldsymbol{\mathcal{G}}}(H)$ will also by satisfied w.e.h.p. by the Erdős-Rényi random graph $\mathbf{G}_{n,m}$ with $m/n \to c/2$.*

2. *If we are impose the additional assumptions:*

   (a) the maximum degree is at most $\epsilon \ln n$ for fixed $\epsilon > 0$,

   (b) the first moment $M_k(\mu_H)$ converges to $M_k(\pi_c) = c^k$,

   then any property which is satisfied w.p.h.p. by $\hat{\boldsymbol{\mathcal{G}}}(H)$ will also hold w.p.h.p. for $\mathbf{G}_{n,m}$ with $m/n \to c/2$.

3. *In both cases, the high-probability guarantee can be weakened; for example, any property which is satisfied a.a.s. by $\boldsymbol{\mathcal{G}}(H)$ will also hold a.a.s. for $\mathbf{G}_{n,m}$.*

3.5.2 Power Law Distributions

Power-law distributions have received considerable attention recently, as it has been observed that many large graphs which occur in the real world exhibit power-laws (e.g. [30, 47]; see discussion in the introduction). One popular method for generating random power-law graphs, introduced by Aiello, Chung, and Lu [2] is simply to choose uniformly among all graphs with power-law degree distributions using the configuration model.

The use of the configuration model as a model for real-world graphs is problematic for a number of reasons, which were touched upon in the introduction. Perhaps the most significant of these is that random graphs generated by the configuration model are locally tree-like, while real-world power-law graphs exhibit dense clusters and short cycles.

A more subtle problem is that the "power-law" refers to the behavior of a degree distribution $\lambda(i)$ in the limit as $i \to \infty$. However, random graph properties are much more sensitive to the lower ranges of the degree distribution; changes in, say, $\lambda(1)$ and $\lambda(2)$ can have huge repercussions for the structure of a random graph. Accordingly, there is a question as to "which" power-law distribution is the correct one, and the fact that $\lambda(i)$ drops polynomially with i for large values of i does not necessarily indicate that this polynomial relationship will hold exactly for all degrees.

With these caveats in place, it is nevertheless appropriate to mention the model of Aiello, Chung, and Lu, if only for historical purposes. The degree distribution proposed in [2] is a strict power law of the form

$$\lambda_\beta(i) = \frac{i^{-\beta}}{\zeta(\beta)},$$

with the zeta function $\zeta(\beta) = \sum_{n=1}^{\infty} n^{-\beta}$. The residual distribution is therefore given by

$$\mu_\beta(i) = \frac{(i+1)^{1-\beta}}{\zeta(1-\beta)},$$

and the corresponding p.g.f. is

$$\psi_{\mu_\beta}(z) = \frac{\mathrm{Li}(\beta-1, z)}{z\zeta(\beta-1)},$$

with the polylogarithm $\mathrm{Li}(x, z) = \sum_{n=1}^{\infty} \frac{z^n}{n^x}$.

3.5.3 Random Regular Graphs

The earlier applications of the configuration model were to the generation of random regular graphs. In this case, the degree and residual distributions are trivial, since every vertex has the same degree. We briefly note that our convergence assumptions in section 3.3 are not strong enough to ensure that $\boldsymbol{\mathcal{G}}(\mu)$ is a regular graph, even if $\mu(i) = 1$ for some i, for the simple reason that convergence of the residual distribution allows for the possibility of $o(n)$ vertices of some other degree.

Random regular graphs, particularly of very low degree (i.e. 3, 4, 5) provide a sort of testing ground for algorithmic methods, in that the analysis of such algorithms is greatly simplified by the fact that the parametrization is trivial. As such it is often possible to explore the limits of the algorithmic method more fully, dealing with infinite dimensionality, etc. In chapter 11, we will exploit this ability by examining a heuristic for the independent set problem on 3-regular graphs, the analysis of which would require significantly more effort for the general random graph $\boldsymbol{\mathcal{G}}(\mu)$, or even for the Erdős-Rényi random graph $\mathbf{G}_{n,p}$.

Part II

Methods

Chapter 4

Topological Representation and Differential Equations

In the previous chapter, we introduced the basic objects which we shall be studying, the kinds of questions we shall be asking, and the general strategy for finding answers. Modulo certain technicalities, the essential facts are:

- the objects we are interested are *asymptotic random graphs* $\boldsymbol{\mathcal{G}}(\mu)$, which are sequences of $\mathbf{G}_\eta$ of random graphs (more precisely, configurations), such that the *residual distribution* converges to μ.
- we wish to determine which graph properties are satisfied *with high probability (w.h.p.)* by $\mathbf{G}_\eta$ in the limit $n \to \infty$;
- our main approach is the *algorithmic method*, which involves predicting the output of an algorithm which computes a given property.

In this chapter, we address the next natural question: how do we plan to predict the outcome of algorithms for which the input is a random graph $\boldsymbol{\mathcal{G}}(\mu)$?

The short answer is: by solving differential equations. Using a technique developed by Wormald [65], the sample path of certain discrete random processes can be predicted with asymptotic precision via differential equations.

A typical invocation of the differential equations method can be described in three phases. The first phase is where we devise an algorithm which computes our graph property, and compute the transition probabilities and other basic characteristics of the corresponding execution path (or, more precisely, the *observable process* as defined in §3.2.2). The nature of this first phase is largely combinatorial and probabilistic: we are dealing with edges, vertices, conditional probabilities, and so on.

Let us skip the second phase for now and discuss the third phase, which involves solving differential equations. At this point, we are working with a problem which is entirely different in nature. The basic objects in play are continuous functions, derivatives, etc., and there is no longer any "randomness," "discreteness," or "asymptoticness." The third phase is just a differential equations problem.

The second phase, which includes everything that takes place in between, is the subject of the current chapter. Depending on what we consider "difficult," the second phase is either the easiest or the most difficult of the three. In one sense, the second phase is "trivial." Devising an algorithm and solving a system of differential equations both involve "thinking" and "work," while translating the description of a random process to a system of differential equations is, at least in principle, a fairly routine and automated task.

On the other hand, the first and third phases involve solving relatively well-defined problems, each of which is confined to a well-understood and coherent mathematical domain. The second phase requires us to cross these boundaries, and, while the way to do this is indeed somewhat apparent, there are several technicalities which must be addressed. The correct adjective is perhaps "unpleasant" rather than "difficult."

Our present objective is to put in place the underlying machinery which will make the transition from the discrete, probabilistic realm of random graph algorithms, to the continuous realm of differential equations, as effortless as possible. An ideal outcome would perhaps be a "black-box" theorem: the assumptions would be conditions about a random process, and the conclusions would be a system of differential equations, the solution of which would determine, with high probability, the sample path (or at least the terminal state) of the given random process.

Unfortunately, the situation is not so simple. The problem is not in proving such a "master differential equations theorem," since there is nothing particularly difficult or surprising about the fact that discrete random processes behave similarly to differential equations when scaled appropriately. The problem is constructing a sufficiently robust theorem statement; indeed, the proof of such a "master theorem" would likely be easier and

perhaps even shorter than than the statement itself.

While it is possible to state an acceptably general theorem (e.g. the main theorem in [65]), the differential equations method is ultimately a "method," and not a "theorem." The guiding philosophy of this method, as we shall practice it, is that once we have created an algorithm and determined the basic characteristics of its execution path, it should no longer be necessary to explicitly deal with combinatorial structures, probabilities, or asymptotic sequences. Instead, we should be able to complete the analysis of this algorithm by manipulating derivatives, continuous functions, and other such objects.

Chapter Organization

We begin in section 4.1 by giving a slightly more detailed overview of our approach to the differential equations method. Once these basic concepts in are introduced, much of the chapter will become somewhat predictable, and a reader comfortable with the fact that all i's can been dotted and t's crossed can rely mainly on this first section for a general understanding.

A more detailed preview of the remaining sections will be given in §4.1.3, at the end of section 4.1. Briefly, in sections 4.2 and 4.3 we develop two layers of abstraction which will ease the translation from discrete random objects to continuous function spaces. The first involves logical inference among asymptotic events, and the second involves topological representation of discrete random structures.

In section 4.4 we discuss specific issues related to the topological representation of random processes in function spaces. Finally, in section 4.5, we prove a version of the differential equations theorem of Wormald [65], which, due to the framework which will be in place by then, is an almost immediate consequence of Azuma's martingale concentration inequality.

History and Background

The presentation of the differential equations method in this chapter differs substantially from the original version in [65], in that we emphasize logical and topological abstraction rather than actual differential equations. Also, the differential equations theorem in section 4.5 is proved by a different technique; in fact, as noted above, this theorem hardly requires proof once the requisite foundations are in place.

As such, the material in this chapter draws more from sources outside of probabilistic combinatorics. The logical abstraction in section 4.2 is motivated by ideas in non-standard analysis, lattice completions, and representation theorems for Boolean algebras. We deal with these subjects at a relatively elementary level, which does not remotely approach the depth of, say, Stone's representation theorem. An in depth discussion of lattice theory, the relationship of logic to topology, and other such topics can be found in various text, such as [17, 41, 64].

The technical substance of section 4.3 is largely a review of elementary point set-topology: properties of Compact-Hausdorff spaces, the Arzelà-Ascoli theorem, etc. Similarly, the technical substance of section 4.4 is a review of elementary calculus: the Riemann-Stieltjes integral, functions of bounded variation, etc. These classical results about compactness and convergence in function spaces form the substance of the proof of the differential equations theorem in section 4.5, which becomes more or less a corollary to martingale concentration.

Both the logical/topological abstraction and the differential equations techniques can clearly be developed far beyond what appears in this chapter. Our stopping point is based, perhaps somewhat arbitrarily, on what is needed for the applications in this dissertation. As readers familiar with either of these areas will recognize, the path to further abstraction is relatively clear in both cases, and the question is whether these additional abstractions will prove useful.

4.1 Overview of Methodology

The differential equations method can be understood as a special instance of a more general approach to reasoning asymptotically about discrete random structures using basic ideas from point-set topology: compactness, convergence, and so on. Our overview thus begins in §4.1.1 by discussing this topological abstraction, without reference to differential equations.

In applications, though, this topological abstraction will give way to actual function spaces and differential equations. In order to illustrate how this works, we will walk through a simple example in §4.1.2. This will also illustrate our basic use of notation and terminology, to the point that the reader should be able to understand the use of such notation in subsequent chapters without much difficulty.

The direction this chapter is headed should become relatively apparent based on the abstract discussion in §4.1.1, and the more concrete example in §4.1.2. We conclude this section in §4.1.3 by going over what must be done in order to make things work out rigorously.

4.1.1 The Topological Method

The essential concept in this methodology is *topological representation*; in the abstract, a topological representation is simply an asymptotic sequence $\mathbf{x}_\eta$ of random elements in a topological space X, called the *representation space*. The asymptotic random element $\mathbf{x}_\eta$ is meant to represent "all" of the information required to solve whatever random graph problem we are interested in.

One important requirement is therefore that, if we are able to show that $\mathbf{x}_\eta \to x$ w.h.p. for some fixed $x \in X$, then our problem will be "solved." This is not difficult to accomplish, since we may express graph properties as mappings α from the set $\mathbb{G}$ of graphs to some (possibly trivial) topological space. The topological representation, though, will not be limited to the image $\alpha(\mathbf{G}_\eta)$; the representation space X will also include many other

parameters which, while not of direct interest, are necessary in order to determine the w.h.p. limit of the "important" parameter $\alpha(\mathbf{G}_\eta)$.

Since our "goal" is to find a fixed $x \in X$ to which $\mathbf{x}_\eta$ converges w.h.p., we will call this a *solution*. More precisely, we shall define the following:

- if $\mathbf{x}_\eta \nrightarrow x$ w.h.p. then x is a *w.h.p. non-solution*, and otherwise x is a *w.h.p. weak solution*;
- if $\mathbf{x}_\eta \to x$ w.h.p. then x is a *w.h.p. strong solution*.

The expression "$\mathbf{x}_\eta \nrightarrow x$", we recall, is a (w.h.p.) *separation*, and specifies that there exists a fixed open $B \ni x$ such that $\mathbf{x}_\eta \notin B$ w.h.p., so the symbol "$\nrightarrow$" stands for "is bounded away from" rather than "does not converge to." A weak solution is thus a point x such that "$\mathbf{x}_\eta$ is not bounded away from x w.h.p.",[1] which is roughly analogous to an accumulation point for an infinite non-random sequence, while a strong solution is an actual w.h.p. limit.

The reason for renaming these concepts derives from the way we intend to use them. One way to go about finding a strong solution (the naive way, perhaps), is to use what we know about $\mathbf{x}_\eta$ to explicitly prove that, for any $B \ni x$, the sequence $\mathsf{P}[\mathbf{x}_\eta \in B]$ tends to 1 sufficiently quickly as $n \to \infty$. In this case, the topology of the space X serves primarily as a language in which to describe our problem, and our only real interaction with this topology takes place when we choose an arbitrary $B \ni x$ with the intention of proving that $\mathbf{x}_\eta \in B$ w.h.p.

We intend to use the representation space as a tool for actually finding a solution. This entails translating our knowledge of the problem at hand into the language of the topology of X, and we shall use the "constraint satisfaction" metaphor for this purpose:

- a *w.h.p. solution constraint* is a condition that is satisfied by every w.h.p. weak solution.

[1] This should be parsed as "it is not the case that ($\mathbf{x}_\eta$ is bounded away from x w.h.p.)".

There are many ways to derive solution constraints, and much of this chapter is devoted to describing these.

The important thing about solution constraints is that they are constraints on *possible solutions*, so they can be stated without reference to random structures or asymptotic sequences or anything else besides the topology on X. So, once we have converted our knowledge about $\mathbf{x}_\eta$ into solution constraints, it is no longer necessary to reference any of these, and even the topological representation $\mathbf{x}_\eta$ itself becomes superfluous.

The idea is to use these solution constraints, along with the topological structure of X, to rule out all possible weak solutions except for one, and then conclude that this unique solution is in fact a strong solution. In order to draw this conclusion, what is required is compactness.

If the representation space X is compact, then by choosing finite sub-covers as usual, a unique weak solution can be converted to a strong solution; this is for the same reason that a non-random sequence in a compact space must converge to a unique accumulation point. Typically, we will invoke a weaker condition which achieves the same result:

- a *w.h.p. solution space* is a compact subset X_0 such that $\mathbf{x}_\eta \to X_0$ w.h.p.[2]

The existence of a compact solution space is sufficient to ensure that a unique weak solution must be a strong solution, and in this way, compactness validates the "constraint satisfaction" metaphor.[3]

At a high-level, these four definitions characterize what we may call the "topological method," which is ultimately quite straightforward. First, construct a topological representation $\mathbf{x}_\eta$ of the relevant discrete structures in a space X which is either compact or such that there is some compact solution space. Then, express our knowledge in terms of

[2]This w.h.p. convergence $\mathbf{x}_\eta \to X_0$ simply means that for any open $B \supseteq X_0$, the event $\mathbf{x}_\eta \in B$ occurs w.h.p.

[3]In to make the metaphor "really" work out, a Hausdorff assumption is also necessary, but this is trivially satisfied in most cases.

solution constraints of the form "every weak solution must satisfy this condition." Finally, show that only one element of X satisfies all of these constraints... "et voilà."

The Differential Equations Method

In practice, the "topological method" described above will in effect turn into a "differential equations method," meaning that the space X will be a function space, the topological representation will involve scaling the sample path of a discrete random process, and the solution constraints will be differential equations.

The topological abstraction plays a role analogous to a "wrapper" in software architecture: while the differential equations do all of the "work," these topological concepts facilitate communication with other parts of an argument, and encourage modularity by keeping implementation details hidden from view. In our situation, the role of wrapper is not at all insignificant.

For starters, the topological approach gives us a language in which to express such a differential equations theorem. Rather than attempting a "silver bullet" theorem statement which goes from combinatorial structures to differential equations and back again, our version will simply give conditions under which a differential solution constraint will hold w.h.p. As we shall see, this will make various technicalities largely irrelevant, including:

- the (possibly infinite) number of dimensions;
- whether the system of differential equations is uniquely solvable;
- "imperfections" in a process definition, e.g. stopping times, assumptions which only hold temporarily or conditionally, and other non-homogeneous behavior;
- details regarding scaling and asymptotic parametrization.

None of these would individually (or even collectively, perhaps) preclude a general theorem statement, but as a whole, technicalities like this present somewhat of a non-trivial annoyance.

However, these problems do not complicate the proof of the essential fact that random processes behave like (solutions to) differential equations; they just make stating this fact more difficult. The topological abstraction will allow us to more or less state that we will cross each such bridge when we come to it.

Moreover, if we encounter some "thing" that is not naturally expressed as a function or a differential equation, it is often quite easy to represent this thing topologically, alongside a general argument which involves differential equations, and perhaps other topological structures as well. As such, this open-ended topological methodology is perhaps the "heart of the matter," while differential equations just happen to be the most powerful applications (that we know of).

4.1.2 A Simple Example

We now illustrate, by example, how this method will work in practice, along with general notation, etc. Hence, consider the following simple problem:

- We are given a set of n coins, each of which is initially tails, and at each of n time steps, we choose a coin uniformly at random and turn it over. At the end, how many of these coins are heads?

Asymptotic Problem Definition

Before getting into differential equations, we need to state the problem more precisely. Let us denote by S our set of n coins, and for each $0 \leq t \leq T = n$, we let $\mathbf{R}_t$ denote the subset of S consisting of coins which are heads at time t. The initial state is $\mathbf{R}_0 = \varnothing$, and for each $1 \leq t \leq T$, we choose a uniformly random element from S, and either add or remove this element from $\mathbf{R}_{t-1}$ to yield $\mathbf{R}_t$. The entire sample path is denoted by $(\mathbf{R}_t)_{t=0}^{T}$ and abbreviated by $(\mathbf{R}_t)$.

Since we will be reasoning asymptotically, each value of n technically gets its "own process," which means that all of the objects defined above should have an extra subscript

"n".[4] Our basic asymptotic sequence of random structures is thus the sequence of sample paths

$$\left((\mathbf{R}_{t,n})_{t=0}^{T_n}\right)_{n=0}^{\infty} \tag{4.1}$$

where each $\mathbf{R}_{t,n} \subseteq S_n$ and $|S_n| = T_n = n$.

Nevertheless, (4.1) marks the first and last appearance of this kind of expression, because from now on, we will drop the implicit asymptotic subscript "n" and return to the notation from two paragraphs above. As usual, when using this convention, all unqualified references to convergence or uses of either the convergence arrow "$\to$" or "big O" notation refer to the limit as $n \to \infty$.

The terminal number of heads $|\mathbf{R}_T|$ is a random integer, which cannot meaningfully converge unless it is scaled. There is more than one way to proceed from here, but our standard way of scaling is as follows.

- We wish to find a fixed $z \in [0,1]$ such that the fraction $|\mathbf{R}_T|/n$ of heads at the end of the process converges to z with probability which is exponentially high with respect to n (n-w.e.h.p.).

Topological Representation

The random structure we are dealing with is the sample path $(\mathbf{R}_t)$, and in order to represent this topologically, both the time and space coordinates must be scaled. There is really only one option for scaling the state space, which is to divide by n; hence, we let $\mathbf{z}_t = |\mathbf{R}_t|/n$ denote the fraction of heads at time t, and we now have a random process $(\mathbf{z}_t)$ in with state space $[0,1]$.

There is perhaps more than one possibility for time-scaling, but the idea is to again divide by $T = n$; we shall use the least integer function $\lfloor x \rfloor = \max\{t \in \mathbb{Z} : t \leq x\}$ to fill in

[4] In fact, even more technically, according to our conventions, even "n" should have a subscript n_η since the symbol η and not n is used to index asymptotic sequences, but we will overlook this for the present example.

the gaps, and represent $(\mathbf{z}_t)$ as a mapping $[0,1] \to [0,1]$ by

$$\xi \mapsto \mathbf{z}_{\lfloor \xi T \rfloor}.$$

We will simplify this notation by simply writing $\mathbf{z}_\xi = \mathbf{z}_{\lfloor \xi T \rfloor}$, which does involve abuse of notation, but should not cause ambiguity for the simple reason that, depending on whether the subscript is an integer t or a real $\xi \in [0,1]$, there is only one possible meaning.[5] We use "tuple" notation $(\mathbf{z}_\xi)_{\xi \in [0,1]}$ to manipulate this mapping as one object, and again abbreviate by $(\mathbf{z}_\xi)$.

The representation space in this case is the space $\mathcal{M}([0,1],[0,1])$ of mappings $[0,1] \to [0,1]$; predictably, we denote fixed elements in this space using the same notation $(z_\xi) = (z_\xi)_{\xi \in [0,1]}$. This space has more than one possible topology, but, by default, we shall use the product topology, which is the topology of pointwise convergence. Hence, a w.e.h.p. solution is a function (z_ξ) such that

$$\mathbf{z}_\xi = \mathbf{z}_{\lfloor \xi T \rfloor} \to z_\xi \text{ w.e.h.p.}$$

for every fixed $\xi \in [0,1]$, and we shall denote this pointwise convergence simply by $(\mathbf{z}_\xi) \to (z_\xi)$.

The other natural choice would be the uniform topology, and since pointwise convergence is the default, we shall denote uniform convergence explicitly by $(\mathbf{z}_\xi) \xrightarrow{\infty} (z_\xi)$. Briefly, there are two main reasons for choosing the product topology: first, $\mathcal{M}([0,1],[0,1])$ is compact under the product topology, and second, pointwise convergence is easier to establish than uniform convergence. There are also arguments in favor of the uniform topology, but, for most processes we encounter, including this one, pointwise and uniform convergence will be equivalent, so ultimately it makes no difference.

[5] Unless $\xi = 0$ or 1, but in this case, we trust that the intended meaning can be discerned without controversy.

The Solution Space

The next general step is to find a *compact solution space*, which is a compact subset of the representation space to which w.e.h.p. convergence can be established. In this case, the space $\mathcal{M}([0,1],[0,1])$ is already compact under the product topology, so this step is not strictly necessary. Nevertheless, general elements of the product space $\mathcal{M}([0,1],[0,1]) \simeq [0,1]^{[0,1]}$ are "functions" in name only. In practice, these are just "tuples," as well they should be, since this more or less the purpose of the product topology. These "tuples" are not differentiable, or continuous, or anything else — lack of continuity does not begin to describe the horrible things that can occur in the product space $\mathcal{M}([0,1],[0,1])$.

We will thus find a restricted solution space, not for compactness reasons, but so that we can work with well-behaved functions. Returning to the process definition, since at most one coin changes each step, then the discrete process $(\mathbf{z}_t)$ will satisfy a Lipschitz condition $|\Delta \mathbf{z}_t| = |\mathbf{z}_t - \mathbf{z}_{t-1}| \leq 1/n$. The analogous condition is easily shown to hold for any possible solution, so we have a solution space

$$\mathcal{Z} = \{(z_\xi) \in [0,1]^{[0,1]} : |z_\xi - z_\zeta| \leq |\xi - \zeta| \text{ for all } \xi, \zeta \in [0,1]\} \tag{4.2}$$

This is a much more friendly space; for starters $\mathcal{Z}$ is a subset of the space $\mathcal{C}([0,1],[0,1])$ of continuous functions. The Lipschitz condition also ensures that these functions have bounded variation, are differential almost everywhere, and are generally well-behaved.

Another important consequence of the Lipschitz condition is that pointwise convergence implies uniform convergence (and hence the two are equivalent); and, for similar reasons, the uniform and product topologies coincide on the solution space $\mathcal{Z}$, which is therefore compact under this topology. As promised, the Lipschitz condition makes it unnecessary to choose between the two.

In section 4.2 we shall introduce definitions which allow us to express the fact that pointwise implies uniform convergence (and other similar things) by

$$(\mathbf{z}_\xi) \to (z_\xi) \implies (\mathbf{z}_\xi) \xrightarrow{\infty} (z_\xi) \text{ w.e.h.p.}$$

In our case, the above implication is not related to exponentially high probability, so really it holds *ultimately always (u.a.)* rather than just w.e.h.p.

We finally point out that, in an important sense, restricting the solution space to $\mathcal{Z}$ really does have to do with compactness. We were able to avoid this issue by using the product topology by default, but this is a form of "cheating," because the correct topology on the space of "functions" and not "tuples" is the uniform topology.

Solution Constraints

We are now at the point where we can derive *differential solution constraints*. For this purpose, we note that the expected increments of our discrete random process can be computed rather easily: since the probability of choosing a coin which is a head in step $t+1$ is precisely $\mathbf{z}_t$, we have

$$\mathsf{E}_t[\Delta \mathbf{z}_{t+1}] = \frac{(1-\mathbf{z}_t) - \mathbf{z}_t}{n} = \frac{1-2\mathbf{z}_t}{n}. \tag{4.3}$$

The differential equations theorem (theorem 4.5.3) states, in effect, that this equation, along with the Lipschitz condition, yields the solution constraint:

$$\text{every w.e.h.p. (weak) solution must satisfy} \quad z_\xi = \int_{\zeta=0}^{\xi} (1-2z_\zeta)d\zeta \quad \text{for all } \xi \in [0,1]. \tag{4.4}$$

This is of course an integral rather than differential equation, and the distinction is not completely trivial since integral equations can be satisfied by functions which are not differentiable.

Regardless, we will often write integral constraints in differential form, technicalities notwithstanding. And, rather than writing "every solution..." we will simply write:

$$\text{the constraint } dz_\xi = (1-2z_\xi)d\xi \text{ holds w.e.h.p.} \tag{4.5}$$

In the present case, any solution to (4.4) must of course be differentiable, so it makes no difference.

Note that the high-probability guarantee "w.e.h.p." in these constraints refers to the function (z_ξ) which is not an asymptotic random structure, but a fixed element of a

function space; this should not be controversial in light of our general discussion about what a "w.e.h.p. solution" is.

Finding a Solution

At this point, we simply have an initial value problem to solve. The differential equation $dz_\xi = (1 - 2z_\xi)d\xi$ with $z_0 = 0$ has unique solution

$$z_\xi = \frac{1 - e^{-2\xi}}{2}.$$

Using compactness, we may thus conclude that this is a w.e.h.p. strong solution, and this solves our problem: the fraction of heads at the end of the process converges, with exponentially high probability, to the value $(1 - e^{-2})/2$.

At a more granular level, the actual argument is:

1. the topological representation $(\mathbf{z}_\xi)$ is a (sequence of) random element(s) in the compact product space $\mathcal{M}([0, 1], [0, 1])$;

2. the discrete Lipschitz condition has two implications:

   - the set $\mathcal{Z}$ is solution space, so $(\mathbf{z}_\xi) \nrightarrow (z_\xi)$ w.e.h.p. for any $(z_\xi) \notin \mathcal{Z}$;
   - pointwise convergence and uniform convergence are now equivalent;

3. the formula (4.3) for expected increments (and Lipschitz condition, again) implies that integral constraint (4.4) holds w.e.h.p.;

4. the unique w.e.h.p. weak solution is $z_\xi = \frac{1-e^{-2\xi}}{2}$, and by compactness this is a strong solution;

5. this implies that $\mathbf{z}_T \to (1 - e^{-2})/2$ w.e.h.p.

In this particular case, differential equations are not really necessary (not to mention "topological representation"), and the solution can be determined — or at least guessed correctly — by summing the odd values of the Poisson distribution:

$$z_\xi = \sum_{i=0}^{\infty} \frac{e^{-\xi}\xi^{2i+1}}{(2i+1)!} = e^{-\xi}\sinh(\xi) = \frac{1 - e^{-2\xi}}{2}.$$

On the other hand, once the groundwork is in place, it is not necessary to go through all of these steps explicitly, and the differential equations method will imply almost immediately, not only that this is the correct guess, but also that $(\mathbf{z}_\xi) \to (z_\xi)$ w.e.h.p.

4.1.3 Organization of the Chapter

The rest of this chapter consists primarily of two kinds of material:

- "bookkeeping" which allows this kind of argument to run smoothly;
- "bells and whistles" which handle certain common variations.

The overview given above is sufficiently detailed that the bookkeeping should be somewhat predictable, with at least one exception: an additional layer of abstraction, called *w.h.p. logic*, resides between actual random structures and their topological representation. The entire progression from discrete random structures to topological reasoning in a solution space is shown in figure 4.1.

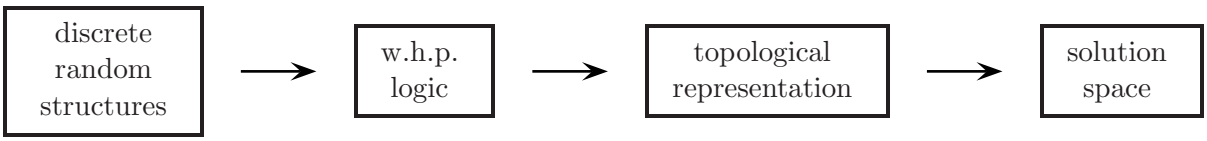


Figure 4.1: Layers of abstraction for the topological/differential equations method.

In the first stage, we are dealing with random discrete objects, probabilities, asymptotic sequences, and so on. In the second stage, the relevant objects are asymptotic events, and all that matters about an asymptotic event is whether or not it occurs w.h.p. There are no more discrete objects or probabilities, only a big "pile" of logical entities, some of which occur w.h.p. and some of which do not.

In the third stage, asymptotic events are organized topologically into convergences and separations. Now we are dealing with the topological representation, which is a random element in a topological space, and the only question is where (and if) it converges w.h.p. In

the final stage, all that remains is the solution space, which is the set of possible w.h.p. limit points, along with solution constraints, which are statements of the form "every solution must satisfy this condition."

This progression is described in the next two sections: in section 4.2 we transition from discrete random structures to w.h.p. logic, and in section 4.3 we move through topological representation to the solution space.

Once this development is complete, we begin to add bells and whistles. In the context of the algorithmic method, the most important random structure is the discrete random process, and in section 4.4, we describe how various attributes of a discrete random process will be represented topologically. The general idea is simply to scale the state space and the time coordinate as described above, but since a random process is not actually a continuous function, there are other considerations, such as random times, time-dependent events, and so on.

The "differential equations" portion of the chapter is divided between sections 4.4 and 4.5. The differential equations method is based on two simple ideas. The first is that the increments of a discrete random process, when scaled appropriately, resemble the derivative of a continuous function. The second is that the sample path of certain discrete processes can be well-approximated by summing expected increments rather than true increments.

Neither of these observations are difficult to make rigorous. Modulo minor technicalities regarding jump-discontinuities convergence of the Riemann and Riemann-Stieltjes integrals is immediate in most cases due to uniform convergence. And, if a Lipschitz condition holds, the approximability of a process by expected increments is a direct consequence of martingale concentration. Hence, although differential equations are perhaps the substance of this entire methodology, there is not much left to prove once everything is correctly defined. In a sense, martingale concentration will simply "plug in" to the larger topological framework, and the rest of the work is taken care of by structural properties of function spaces.

At the end of all this we will be left with, not a single "silver bullet" theorem statement, but a collection of techniques for deriving solution constraints based on the specification of a discrete random process. We conclude the chapter by giving some insight into how these techniques will work together as a cohesive methodology.

The path of abstraction described above might seem lengthy, but in applications, the two middle layers will largely disappear: most of the reasoning will take place either in the combinatorial realm of graphs, algorithms, and probabilities, or in the analytic realm of continuous functions and differential equations. As such, many of the definitions from the next two sections, where these intermediate layers are developed, will not be used much beyond this chapter. In fact, the concepts of "w.h.p. logic" from section 4.2 serve primarily as a platform for topological representation, and as such they will rarely appear, in explicit form, beyond section 4.3 in this same chapter.

Nevertheless, the fact that these layers are there in the background provides a buffer against unforeseen technicalities. If something comes up which does not literally translate to a solution constraint using the results of this chapter, we can temporarily reach back to the topological representation layer; if this is still not sufficient, we can go back to w.h.p. logic, which is the least "user-friendly," but the most robust of the layers.

4.2 High Probability Logic

High-probability logic or "w.h.p. logic," briefly defined, is logic inference in a setting where the basic entities are asymptotic sequences of events, and the truth condition is occurrence with high-probability. In the larger context of the differential equations method, w.h.p. logic represents the first in the series of abstractions described in §4.1.3.

As a setting for this first abstraction, we will assume that we are given a "set of all" asymptotic events; in applications, these events will express properties of certain discrete random structures, but for now, where they come from does not matter. Moreover, although each asymptotic event L_η is in fact an infinite sequence $(L_1, L_2, \ldots)$, we will treat them as individual objects: we just call them "events," and denote them with unsubscripted symbols

as in "L."

The important facts about asymptotic events are the following.

1. The set of all asymptotic events is a Boolean algebra, which we denote by $\mathcal{B}$, and in which the lattice operations are performed pointwise in the obvious way: $L \wedge K$ denotes the sequence $(L_1 \wedge K_1, L_2 \wedge K_2, \ldots)$, and so on.

2. The only relevant question about any $L \in \mathcal{B}$, is whether this event occurs w.h.p. While this depends on the asymptotic behavior of $\mathsf{P}[L_\eta]$ a $\eta \to \infty$, in the abstract, this is just a mapping $\mathsf{whp} : \mathcal{B} \to \{\mathsf{true}, \mathsf{false}\}$ which we call the *w.h.p. valuation.*

3. The w.h.p. valuation satisfies some basic rules, the most significant of which is: if both $\mathsf{whp}(L)$ and $\mathsf{whp}(K)$ are true, then so is $\mathsf{whp}(L \wedge K)$. As a consequence, the set of all events which occur w.h.p. is a *filter* in the Boolean algebra $\mathcal{B}$.

What we are calling "w.h.p. logic" involves lattice operations on a Boolean algebra, and is thus not literally "logic," which deals with formal symbolic expressions. As such, the "rules" of w.h.p. logic, which we describe in this section, are really just restatements of certain basic facts about lattices and quotient structures. Moreover, this lattice-theoretic language serves only to formalize certain semantic constructions; in applications, we will not discuss the "Boolean algebra of asymptotic events," the "w.h.p. filter," or other such objects explicitly. Instead, these objects will be manipulated implicitly, either using topological representation, or in the "plain English" of ordinary mathematical discourse.

Hence, before we proceed with formal definitions, we give a brief summary, which, while not strictly "rigorous," should be enough to convince the reader that the ensuing formalities are both possible and fairly routine.

Brief Summary

There are two key developments in this section. The first involves expressing topological ideas logically (or vice-versa) in terms of the set of all w.h.p. events (i.e. the w.h.p. filter).

- A *meta-event* is something that can be deemed to occur w.h.p., but is not necessarily an event.
- A standard example is a *convergence* $\mathbf{x} \to x$, which is valuated by $\forall_{B \ni x} \mathsf{whp}(\mathbf{x} \in B)$.[6]
- A *meta-implication* has the form $\mathbf{x} \to x \implies \mathbf{y} \to y$, and is valuated by

$$\forall_{B_y \ni y} \exists_{B_x \ni x} \mathsf{whp}(\mathbf{x} \notin B_x \vee \mathbf{y} \in B_y).$$

- Other meta-event constructions are possible, their meaning is apparent, and they can all be reduced to a meta-implication; e.g. a separation $\mathbf{x} \nrightarrow x$ is equivalent to $\mathbf{x} \to x \implies \bot$.
- If $\mathcal{L}$ is an event or a convergence (but not a meta-implication), the *relative w.h.p. valuation*

$$\mathsf{whp}_{\mathcal{L}}(K) = \mathsf{whp}(\mathcal{L} \implies K)$$

  satisfies the same axioms as the original w.h.p. valuation, so we can "assume" that $\mathcal{L}$ occurs w.h.p., and the same "rules" will apply.
- These meta-events can be expressed in "plain English" as in "if $\mathbf{x} \to x$ then $\mathbf{y} \to y$ w.h.p."

The second development has to do with compactness.

- For any $X_0 \subseteq X$ the convergence $\mathbf{x} \to X_0$ can be defined naturally by $\forall_{B \supseteq X_0} \mathsf{whp}(\mathbf{x} \in B)$.
- If X_0 is compact, and L is an event, then the following are equivalent:
  1. $\mathsf{whp}(\mathbf{x} \to X_0 \implies L) = \mathsf{true}$,
  2. for all $x \in X_0$, $\mathsf{whp}(\mathbf{x} \to x \implies L) = \mathsf{true}$.

[6] Here $\mathbf{x}$ is actually an asymptotic sequence $\mathbf{x}_\eta$ with the subscript omitted. In the current context, it is appropriate to think of a convergence as a logical statement about events of the form $(\mathbf{x} \in B)$ rather than a sequence which tends to a limit.

- The same is true for any meta-event $\mathcal{L}$, and the intuitive meaning is: anything that can be proved under the assumption that $\mathbf{x} \to x$ for a fixed but unknown element $x \in X_0$ can also be proved under the (nominally weaker) assumption that $\mathbf{x} \to X_0$.
- In this situation, if $\mathsf{whp}(\mathbf{x} \nrightarrow x) = \mathsf{true}$ for every $x \neq x_0$, then $\mathsf{whp}(\mathbf{x} \to x_0) = \mathsf{true}$.

4.2.1 Filters in Boolean Algebras

High-probability logic takes place in the Boolean algebra of asymptotic events, and thus we shall make use of some elementary concepts from lattice theory. We will not attempt to give a comprehensive introduction to this topic, and instead present a brief but self-contained summary of key definitions and facts. The reader is referred to [17] for proofs and additional background.

We assume familiarity with a *Boolean algebra*, which is a distributive lattice $\mathcal{B}$ with maximal and minimal elements $\top, \bot$ and complements $\neg L$. The lattice structure induces a partial order on $\mathcal{B}$ by $L \leq K$ if and only if $L \vee K = K$, and the meet and join are respectively the infimum and supremum with respect to this partial order. Boolean algebras differ from formal systems of logic, but there are obvious similarities between the two: the maximal element $\top$ can be understood as "true," a meet $L \wedge K$ as a conjunction, and so on. The partial order $L \leq K$ then loosely corresponds to logical implication.

The meet and join in a Boolean algebra are binary operations, and it follows by associativity that all finite subsets of $\mathcal{B}$ have a meet and join, but this is not necessarily true for infinite subsets. If every subset has a meet and a join, then $\mathcal{B}$ is *complete*. Although the Boolean algebras we encounter in applications will indeed be complete, in this section we will only consider finite joins and meets. For the record: we specifically do not assume that $\mathcal{B}$ is complete, so infinite joins and meets are not well-defined, and anything that might appear to be an infinite join or meet is not taking place in this Boolean algebra.

Filters and Valuations

Under the "logic" interpretation of a Boolean algebra, a structure called a *filter* is roughly analogous to a logically coherent set of beliefs.

Definition 4.2.1. For a subset $\mathcal{L} \subseteq \mathcal{B}$:

- $\mathcal{L}$ is *filtered* if $\mathcal{L} \neq \varnothing$ and, for any $L_1, L_2 \in \mathcal{L}$, there exists $L_0 \in \mathcal{L}$ such that $L_0 \leq L_1 \wedge L_2$;
- $\mathcal{L}$ is an *upper set* if, for any $L \in \mathcal{L}$ and $K \in \mathcal{B}$, if $L \leq K$ then $K \in \mathcal{B}$;
- $\mathcal{L}$ is simply a *filter* if both conditions hold.[7]

We will often find it useful to express membership in a filter as follows.

Definition 4.2.2. A *valuation* is a mapping $\mathsf{val} : \mathcal{B} \to \{\mathsf{true}, \mathsf{false}\}$ which satisfies the following:

1. $\mathsf{val}(\top) = \mathsf{true}$;
2. if $\mathsf{val}(L) = \mathsf{true}$ and $L \leq K$ then $\mathsf{val}(L) = \mathsf{true}$;
3. if $\mathsf{val}(L) = \mathsf{val}(K) = \mathsf{true}$ then $\mathsf{val}(L \wedge K) = \mathsf{true}$.

It is easy verify that valuations correspond bijectively to filters, in that any mapping $\mathsf{val} : \mathcal{B} \to \{\mathsf{true}, \mathsf{false}\}$ satisfies the above conditions if and only if the set $\mathsf{val}^{-1}(\mathsf{true})$ is a filter. This definition is thus entirely redundant and exists merely for convenience, and we may of course move freely between filters and valuations to suit are needs. The notation for this is largely self-explanatory; for intuitive reasons, we typically start out with a valuation, in which case the corresponding filter is denoted by $\mathfrak{F}[\mathsf{val}] = \mathsf{val}^{-1}(\mathsf{true})$.

Valuations allow us to directly express the sentiment that "L is true" by $\mathsf{val}(L) = \mathsf{true}$, rather than $L \in \mathfrak{F}[\mathsf{val}]$, which would mean that "L belongs to the set of things that are

[7] Sadly, according to these standard definitions, being *a filter* is is not the same as being *filtered*.

true." The "soundness" of this logical interpretation is perhaps somewhat questionable. On one hand, the "truth" evidently satisfies the three conditions in definition 4.2.2, but, there is at least one loophole: it may be the case that $\mathsf{val}(\bot) = \mathsf{true}$ which suggests that we are prepared to believe "false." In this case, the corresponding filter is the entire algebra $\mathfrak{F}[\mathsf{val}] = \mathcal{B}$; a *proper filter* is a filter which is a proper subset of $\mathcal{B}$ (similarly for a *proper valuation*).

Also, the law of the excluded middle does not hold, meaning that it is possible that $\mathsf{val}(L) = \mathsf{val}(\neg L) = \mathsf{false}$. Accordingly, the condition $\mathsf{val}(L) = \mathsf{true}$ is perhaps best interpreted as meaning "is known to be true." A proper filter which contains either L or $\neg L$ for every $L \in \mathcal{B}$ is called by one of several names: a *prime filter*, a *maximal filter*, or an *ultrafilter*.[8] Ultrafilters are somewhat more complicated objects than ordinary filters, and we will neither make use of them nor discuss them any further.

Induced Filters

One important property of the set of filters in $\mathcal{B}$ is that it is closed under the taking of arbitrary intersections. Briefly, this is because every filter contains $\top$, so the intersection of filters is non-empty, and the other axioms are obviously preserved under intersection. As a consequence, for any subset $\mathcal{L} \subseteq \mathcal{B}$, we may *induce* a filter in the natural way, which is to let $\mathfrak{F}[\mathcal{L}]$ denote the minimal filter which contains $\mathcal{L}$ (i.e. the intersection of all such filters).

The induced filter can be constructed explicitly, and it is easiest for this purpose to assume that the set $\mathcal{L}$ is already filtered, in which case we have

$$\mathfrak{F}[\mathcal{L}] = \{K \in \mathcal{B} : \exists_{L \in \mathcal{L}}(L \leq K)\}. \tag{4.6}$$

Otherwise, it is necessary first consider all finite meets of elements in $\mathcal{L}$ as in

$$\mathfrak{F}[\mathcal{L}] = \{K \in \mathcal{B} : \exists_{L_1,\ldots,L_i \in \mathcal{L}}(L_1 \wedge \cdots \wedge L_i \leq K)\}. \tag{4.7}$$

[8] In fact, none of these terms are typically defined in this way; see [17] for details. In a Boolean algebra, it can be shown that this condition is equivalent to being a maximal (proper) filter, which is also equivalent to being a prime, which means that if $K \vee L \in \mathfrak{A}$ then one of $K, L \in \mathfrak{A}$. The term *ultrafilter* is equivalent to maximal filter, but is sometimes used to specify a maximal filter in a Boolean algebra of subsets.

We will not prove either of these constructions, but they are quite easy (see [17]).

When inducing filters from more than one set, we will abbreviate by $\mathfrak{F}[\mathcal{L},\mathcal{K}] = \mathfrak{F}[\mathcal{L}\cup\mathcal{K}]$, and we note that the inducing of filters is a commutative (and associative) operation, in the sense that

$$\mathfrak{F}[\mathcal{L},\mathfrak{F}[\mathcal{K}]] = \mathfrak{F}[\mathfrak{F}[\mathcal{L}],\mathcal{K}] = \mathfrak{F}[\mathcal{L},\mathcal{K}].$$

In order to facilitate the construction of filters, we will typically work with filtered subsets $\mathcal{L} \subseteq \mathcal{B}$, in which case the induced filter $\mathfrak{F}[\mathcal{L}]$ in (4.6) does not involve finite conjunctions. Also, in order to alleviate potential ambiguity regarding a "filter" versus "a filtered set" we will denote actual filters explicitly by $\mathfrak{F}[\mathcal{L}]$.

In our continuing analogy to logical deduction, the induced filter represents everything that is a logical consequence of the set of assumptions $\mathcal{L}$. The fact that finite (but not infinite) conjunctions are added to $\mathfrak{F}[\mathcal{L}]$ reflects the intuition that logical inference is a finite endeavor: we may choose any finite set of assumptions from $\mathcal{L}$ to construct a proof of K.

Quotient Algebras

A *congruence* on a Boolean algebra is an equivalence relation with respect to which the lattice operations are faithful, meaning that if $L \equiv L'$ and $K \equiv K'$ then $L \wedge K \equiv L' \wedge K'$, and so on. The equivalence classes modulo a congruence form a Boolean algebra called the *quotient algebra*. Congruences and quotient structures are among the basic ideas in algebras, so we omit proofs and details, again referring the reader to [17].

We will construct congruences and quotient algebras first by using valuations as follows. For any valuation val, we define a relation $K \leq_{\mathsf{val}} L$, which specifies that

$$\mathsf{val}(K \implies L) = \mathsf{true},$$

where $K \implies L$ denotes the element $(\neg K \vee L)$ in $\mathcal{B}$. We now state, without proof, some basic facts about this relation:

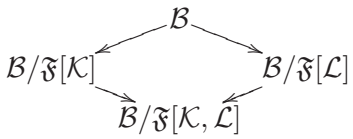
- $\leq_{\mathsf{val}}$ is a pre-order, and the corresponding equivalence relation $\equiv_{\mathsf{val}}$ is a congruence;

- the equivalence classes thus form a Boolean algebra, the *quotient algebra*.

It is a fact that all congruences in a Boolean algebra can be constructed in this way; this is not hard to show (see [17], but it is also not strictly necessary for our present purposes, so we omit a proof.

We denote the quotient algebra by $\mathcal{B}/\mathsf{val}$, and the quotient of any element $K \in \mathcal{B}$ similarly denoted by K/val. Naturally, we may also use a filter in the place of a valuation here, as in $\mathcal{B}/\mathcal{F}[\mathcal{L}]$. The properties of quotient algebras can be summarized informally (yet accurately) by saying that everything works out the way one would expect. We will not elaborate, except to point out:

- the quotient $\mathfrak{F}[\mathcal{L}]/\mathfrak{F}[\mathcal{K}] = \{L/\mathfrak{F}[\mathcal{K}] : L \in \mathcal{L}\}$ of a filter is again a filter in the quotient algebra;
- the operations of taking quotients and inducing filters commute, so $\mathfrak{F}[\mathcal{L}]/\mathfrak{F}[\mathcal{K}] = \mathfrak{F}[\mathcal{L}/\mathfrak{F}[\mathcal{K}]]$;
- quotient-takings commute with each other, as in:



Summary

We now have three ways to express the same logical sentiment:

- $\mathsf{val}(L) = \mathsf{true}$: "$L$ is true"
- $L \in \mathfrak{F}[\mathsf{val}]$: "$L$ is belongs to the set of things that are true"
- $L \equiv_{\mathsf{val}} \top$: "$L$ is equivalent to the truth"

This situation is common in universal algebra, and each of these interpretations serves a purpose. Our definition of a valuation is not standard, but we prefer working with valuations

for intuitive reasons. Filters are often more useful on a technical level, since they can be manipulated set-theoretically. Congruences allow us to construct quotient algebras, which are again Boolean algebras, so everything that holds in our original algebra $\mathcal{B}$ will also hold in any quotient algebra $\mathcal{B}/\mathsf{val}$.

4.2.2 The Algebra of Asymptotic Events

Modulo certain technicalities, it is fairly clear that the set of all asymptotic events corresponding to a given asymptotic sequence of random structures constitute a Boolean algebra.[9] We shall use the same symbol $\mathcal{B}$ to denote this Boolean algebra, in part to emphasize that, for the purposes of w.h.p. logic, an asymptotic event is "just" an element in a Boolean algebra. Similarly, we will continue to denote an asymptotic event by L and call this simply an "event."

Due to the following proposition, we are also able to treat a high-probability guarantee as "just" a valuation.

Proposition 4.2.1. *For any high-probability guarantee defined in §3.4.1, the set of asymptotic events which occur w.h.p. is a proper filter in the algebra of asymptotic events.*

Proof. First, if $L \leq K$, then $\mathsf{P}[L_\eta] \leq \mathsf{P}[K_\eta]$ for all η, and in this case if $\mathsf{whp}(L) = \mathsf{true}$ then so must be $\mathsf{whp}(K) = \mathsf{true}$; hence, the upper-set condition is satisfied. Next, we observe that the high-probability guarantees do not distinguish between rates of convergence of $(1 - \mathsf{P}[L_\eta]) \to 0$ which differ a constant multiple. Since

$$\mathsf{P}[L_\eta \wedge K_\eta] > 1 - \mathsf{P}[\neg L_\eta] - \mathsf{P}[\neg K_\eta],$$

it follows that if $\mathsf{whp}(L) = \mathsf{whp}(K) = \mathsf{true}$, then $\mathsf{whp}(L \wedge K) = \mathsf{true}$.

It thus remains only to show that set of w.h.p. events is neither empty nor equal to all of $\mathcal{B}$, which is immediate since $\mathsf{whp}(\top) = \mathsf{true}$ and $\mathsf{whp}(\bot) = \mathsf{false}$. □

[9] These technicalities, which we will not dwell on, are not related to the lattice structure; this is not at all controversial. The question is whether "all" asymptotic events belong to a set, and this kind of issue has been addressed in §2.1.4 and §**??**.

We shall predictably call $\mathfrak{F}[\mathsf{whp}] = \{L : \mathsf{whp}(L) = \mathsf{true}\}$ the *w.h.p. filter*, and whp is the *w.h.p. valuation*. However, we will develop our methods with respect to a generic valuation val, for which any of the high-probability guarantees, or any other valuation, may be substituted. Hence, except for the fact that elements of $\mathcal{B}$ are now called "events," our basic assumptions and notation are unchanged from the abstract setting above.

At an intuitive level, though, we begin to see why it is useful to distinguish between a valuation and a filter. A valuation represents some "truth" which is "given," while filters can be constructed "on a whim" from a subset $\mathcal{L} \subseteq \mathcal{B}$ of our choosing. These equivalent concepts thus play different roles.

4.2.3 Meta-Events and "W.H.P. Logic"

The lattice theory part of our construction is now more or less complete; in effect, "w.h.p. logic" involves little more than redefining various basic lattice-theoretic concepts in terms of valuations rather than filters and quotient algebras.

For this purpose, we will use first-order symbols $\forall, \exists$ to manipulate valuations, so for a given (possibly infinite) set of valuations $\{\mathsf{val}_i\}_{i \in I}$

$$\mathsf{val}(L) = \forall_{i \in I} \mathsf{val}_i(L) \tag{4.8}$$

is understood to mean that $\mathsf{val}_i(L) = \mathsf{true}$ if and only if $\mathsf{val}_i(L) = \mathsf{true}$ for every $i \in I$. The above expression appears to constitute an "infinite conjunction," but this conjunction does not take place in $\mathcal{B}$, so we have not violated our promise of only using finite lattice operations.

The above infinite conjunction takes place at the "meta-level," and is thus permissible. Moreover, (4.8) does in fact define a valid valuation, due to the fact that the arbitrary intersection of filters is again a filter, and in this case we have $\mathfrak{F}[\mathsf{val}] = \bigcap_{i \in I} \mathfrak{F}[\mathsf{val}_i]$. The logical quantifiers $\forall, \exists$ thus serve essentially as replacements for set-theoretic operations on filters and other subsets of $\mathcal{B}$.

We note, though, the arbitrary union of filters in general not a filter. Hence, while

the expression $\exists_{i\in I}\mathsf{val}_i(L)$ defines a valid mapping from $\mathcal{B}$ to $\{\mathsf{true},\mathsf{false}\}$, this mapping is not in general a valuation.

Definition 4.2.3. For a valuation val and a subset $\mathcal{L}\subseteq\mathcal{B}$, the *relative valuation* $\mathsf{val}_{\mathcal{L}}$ is given by

$$\mathsf{val}_{\mathcal{L}}(K)=\mathsf{true}\ \text{ if and only if }\ K\in\mathfrak{F}[\mathsf{val},\mathcal{L}]. \tag{4.9}$$

As a consequence of the fact that "everything works out," for a single element L, we have

$$\mathsf{val}_L(K)=\mathsf{val}(L\implies K),^{10}$$

and in this case, the relative valuation is just another way to define the pre-order $\leq_L$. Similarly, if $\mathcal{L}$ is a filtered subset of $\mathcal{B}$, then we have

$$\mathsf{val}_{\mathcal{L}}(K)=\exists_{L\in\mathcal{L}}\mathsf{val}(L\implies K)$$

If $\mathcal{L}$ is not filtered, then it is necessary to consider finite conjunctions as in (4.7), but the expression is analogous.

Meta-Events

Given that $\mathsf{val}_L(K)$ is equivalent to $\mathsf{val}(L\implies K)$, for a subset $\mathcal{L}\subseteq\mathcal{B}$, it is natural to think of the condition $\mathsf{val}_{\mathcal{L}}(K)$ as indicating that "$\mathcal{L}$ implies K" under this valuation; the only difficulty is that, while $L\implies K$ is a valid element (i.e. event) in the Boolean algebra $\mathcal{B}$, at this point "$\mathcal{L}\implies K$" is not well-defined.

In order to rectify the situation we will introduce new objects called *meta-events*; simply put, a meta-event is something that is capable of being valuated without belonging to $\mathcal{B}$. Hence, for any valuation val, we have $\mathsf{val}(\mathcal{M})\in\{\mathsf{true},\mathsf{false}\}$ as we would for an ordinary event.

[10] Briefly, if $\mathsf{val}_L(K)=\mathsf{true}$ then clearly $\mathsf{val}(L\implies K)=\mathsf{true}$ since $K\leq(L\implies K)$. Conversely, if $\mathsf{val}(L\implies K)=\mathsf{true}$ then $\mathsf{val}_L(L\implies K)=\mathsf{true}$ also, and clearly $\mathsf{val}_L(L)=\mathsf{true}$ as well, and so $\mathsf{val}_L(L\wedge(L\implies K))=\mathsf{true}$. Since $L\wedge(L\implies K)=L\wedge K\leq K$, this yields $\mathsf{val}_L(K)=\mathsf{true}$.

Now, a valuation val must by definition "know how" to assign any $L \in \mathcal{B}$ a truth-value, but for meta-events, which do not belong to $\mathcal{B}$, this is not part of the job description. Hence, in order to define a meta-event, we must decide how this valuation will be performed. A meta-event can thus be described as a *right-mapping* from the set of valuations to $\{\mathsf{true}, \mathsf{false}\}$: this is a mapping $\mathcal{M} : \mathsf{val} \mapsto \mathsf{val}(\mathcal{M})$, where $\mathcal{M}$ is written on the right-hand side of val. For intuitive purposes, though, we shall think of the valuation as the "function" which does the "valuating."

Of course, we would not want just any right-mapping to qualify as a meta-event; the behavior of the mapping $\mathsf{val} \mapsto \mathsf{val}(\mathcal{M})$ should be "the same" it is for the corresponding mapping $\mathsf{val} \mapsto \mathsf{val}(L)$ for actual events $L \in \mathcal{B}$. The elegant approach here might be to first define what conditions we want a meta-event to satisfy as a right-mapping from valuations to truth-values. However, we shall take a more expeditious route and offer the following constructive definition.

Definition 4.2.4. The following are meta-events:

1. any subset $\mathcal{L} \subseteq \mathcal{B}$ defines a *meta-conjunction*, for which $\mathsf{val}(\mathcal{L}) = \mathsf{true}$ if and only if $\mathcal{L} \subseteq \mathfrak{F}[\mathsf{val}]$;

2. if $\mathcal{L}$ is a meta-conjunction, and $\mathcal{M}$ is any meta-event, then the expression $\mathcal{L} \implies \mathcal{M}$ defines a *meta-implication*, which is valuated by

$$\mathsf{val}(\mathcal{L} \implies \mathcal{M}) = \mathsf{val}_{\mathcal{L}}(\mathcal{M}). \tag{4.10}$$

These valuations can be computed explicitly using set-containment of filters. For a meta-conjunction, we simply have

$$\mathsf{val}(\mathcal{L}) = \forall_{L \in \mathcal{L}} \mathsf{val}(L).$$

For meta-implications, we first consider a case $\mathcal{L} \implies \mathcal{K}$, where both $\mathcal{L}, \mathcal{K}$ are meta-conjunctions. In this case, if $\mathcal{L}$ is filtered, we simply have

$$\mathsf{val}(\mathcal{L} \implies \mathcal{K}) = \forall_{K \in \mathcal{K}} \exists_{L \in \mathcal{L}} \mathsf{val}(L \implies K).$$

As usual, if $\mathcal{L}$ is not filtered, we must replace the set $\mathcal{L}$ in the existential quantification $\exists_{L\in\mathcal{L}}$ with the set $\{L_1 \wedge \cdots \wedge L_i : L_1, \ldots, L_i \in \mathcal{L}\}$ of all finite conjunctions of elements in $\mathcal{L}$.

We now show that the case $\mathcal{L} \implies \mathcal{M}$, where $\mathcal{M}$ is a general meta-event, can be reduced to a meta-implication between two meta-conjunctions.

Proposition 4.2.2. *Let $\mathcal{L}, \mathcal{K}$ be meta-conjunctions, and let $\mathcal{M}$ be a meta-event. Then, for any valuation* val*, we have*

$$\mathsf{val}\Big(\mathcal{L} \implies (\mathcal{K} \implies \mathcal{M})\Big) = \mathsf{val}\Big(\mathcal{L}, \mathcal{K} \implies \mathcal{M}\Big).$$

In particular, every meta-event can be expressed as a meta-implication $\mathcal{L} \implies \mathcal{K}$ between two meta-conjunctions.

Proof. For the main claim, by definition $\mathsf{val}(\mathcal{L} \implies (\mathcal{K} \implies \mathcal{M})) = \mathsf{true}$ if and only if $\mathcal{K} \implies \mathcal{M}$ is true under the valuation $\mathsf{val}_{\mathcal{L}}$, which corresponds to the filter $\mathfrak{F}[\mathsf{val}, \mathcal{L}]$. This in turn is equivalent to the condition that $\mathcal{M}$ is true under the valuation $(\mathsf{val}_{\mathcal{L}})_{\mathcal{K}}$, which corresponds to the filter

$$\mathfrak{F}[\mathfrak{F}[\mathsf{val}, \mathcal{L}], \mathcal{K}] = \mathfrak{F}[\mathsf{val}, \mathcal{L}, \mathcal{K}].$$

This is precisely equivalent to $\mathsf{val}\Big(\mathcal{L}, \mathcal{K} \implies \mathcal{M}\Big)$.

The fact that any meta-event can be reduced to an implication of the form $\mathcal{L} \implies \mathcal{K}$ follows inductively by repeated application of this reduction, along with the fact that any meta-conjunction can be written as $\top \implies \mathcal{L}$. □

Disjunctions and Other Concerns

The above definition gives us a fair amount of expressive ability, but not everything that "might" be a meta-event satisfies this definition, and we now discuss two such cases.

First, definition 4.2.4 does not permit an expression along the lines of

$$\Big((\mathcal{I} \implies \mathcal{J}) \implies (\mathcal{K} \implies \mathcal{L})\Big).$$

At a practical level, this is not a concern, because we will not need the ability to construct such a "meta-event." At a technical level, though, it can be shown that a meta-event which

behaves in the way this expression is intended can be constructed as a meta-implication between to meta-conjunctions. This is not all that difficult, but it is also not entirely trivial, and begins to encroach on the territory of the axiom of choice. In any case, since we do not need this construction, we will leave this stone unturned.

A more pertinent example of something that is not a meta-event is the "disjunction"

$$\text{either } \mathsf{val}(L) = \mathsf{true} \text{ or } \mathsf{val}(K) = \mathsf{true}.$$

Unlike the first example, this is a true "non-meta-event," meaning that a meta-event which behaves this way is impossible to construct (for general L, K) using definition 4.2.4.[11]

The issue is that if we let $K = \neg L$, then the above meta-event would not always be "true," even though the idea it is trying to express — that "either L or $\neg L$" must be true — would seem to be a tautology. Of course, we could allow this to be a meta-event by fiat, but this is not "correct," for reasons both technical and intuitive.

It is possible to express conjunctions as meta-events according to definition 4.2.4, and the way to do this is by

$$\mathcal{L} \implies \bot.$$

Whether this is considered a "meta-disjunction" or a "meta-negation" is largely irrelevant, since it behaves the same way in either case.

If the set $\mathcal{L}$ is filtered, then this meta-event is in fact equivalent to the condition that $\mathsf{val}(\neg L) = \mathsf{true}$ for at least one $L \in \mathcal{L}$. But, if $\mathcal{L}$ is not filtered, then the valuation becomes

$$\exists_{L_1,\dots,L_i \in \mathcal{L}} \Big(\mathsf{val}(\neg L_1 \vee \cdots \vee \neg L_i) = \mathsf{true} \Big).$$

This imposes a restriction on the possible kinds of sets which may be used in disjunctions: if $\mathcal{K}$ is the dual of a filtered set, then the meta-disjunction of $\mathcal{K}$ can be constructed by negating the meta-conjunction $\{\neg K : K \in \mathcal{K}\}$.

[11] Perhaps this statement requires justification, but it is not hard to show; more to the point, the question of whether it is possible to construct such a meta-event is inconsequential.

The dual of a filtered set is called a *directed set*, and it is easy to see that directed disjunctions do not run into problems with the excluded middle as described above, since any directed set that contains both L and $\neg L$ must also contain $\bot$. In applications, though, we will typically not construct such directed disjunctions explicitly, and instead use the above construction $\mathcal{L} \implies \bot$.

4.2.4 Soundness

We now verify that these definitions are sound, in the sense that the logical rules that "should" hold do in fact hold. There are two things to verify: first, that meta-events are well-behaved under the mapping $\mathsf{val} : \mathcal{M} \mapsto \mathsf{val}(\mathcal{M})$, and second, that valuations are well-behaved under the right-mapping $\mathcal{M} : \mathsf{val} \mapsto \mathsf{val}(\mathcal{M})$. The following proposition establishes the necessary soundness conditions in both directions.

Proposition 4.2.3. *Meta-events and valuations satisfy the following:*

1. *if* $\mathfrak{F}[\mathsf{val}_1] \subseteq \mathfrak{F}[\mathsf{val}_2]$ *and* $\mathsf{val}_1[\mathcal{M}] = \mathsf{true}$*, then* $\mathsf{val}_2[\mathcal{M}] = \mathsf{true}$*;*

2. *if* $\{\mathsf{val}_i\}_{i \in I}$ *is a (possibly infinite) set of valuations and* val *is the (well-defined) valuation which satisfies*

$$\mathsf{val}(L) = \forall_{i \in I} \mathsf{val}_i(L),$$

   for all events $L \in \mathcal{B}$*, then* $\mathsf{val}(\mathcal{M}) = \forall_{i \in I} \mathsf{val}_i(\mathcal{M})$ *also holds for any meta-event* $\mathcal{M}$*;*

3. *if* $\mathsf{val}(\mathcal{L} \implies \mathcal{K}) = \mathsf{true}$ *and* $\mathsf{val}(\mathcal{K} \implies \mathcal{M}) = \mathsf{true}$ *then* $\mathsf{val}(\mathcal{L} \implies \mathcal{M}) = \mathsf{true}$*.*

Proof. For the first claim, we may assume $\mathcal{M}$ has the canonical form $(\mathcal{L} \implies \mathcal{K})$, in which case $\mathsf{val}(\mathcal{L} \implies \mathcal{K}) = \mathsf{true}$ if and only if

$$\mathcal{K} \subseteq \mathfrak{F}[\mathsf{val}, \mathcal{L}],$$

and clearly, if $\mathfrak{F}[\mathsf{val}_1] \subseteq \mathfrak{F}[\mathsf{val}_2]$ then we also have $\mathfrak{F}[\mathsf{val}_1, \mathcal{L}] \subseteq \mathfrak{F}[\mathsf{val}_2, \mathcal{L}]$, which implies the claim.

For the second claim, as discussed above, this infinite conjunction is indeed a well-defined valuation, which corresponds to intersection of filters

$$\mathfrak{F}[\mathsf{val}] = \bigcap_{i\in I}\mathfrak{F}[\mathsf{val}_i].$$

Now, consider a canonical form meta-event $\mathcal{M} = (\mathcal{L} \implies \mathcal{K})$, so by definition $\mathsf{val}(\mathcal{M}) = \mathsf{true}$ if and only if $\mathcal{K} \subseteq \mathfrak{F}[\mathsf{val}, \mathcal{L}]$. Similarly, $\forall_{i\in I}\mathsf{val}_i(\mathcal{M}) = \mathsf{true}$ if and only if $\mathcal{K} \subseteq \bigcap_{i\in I}\mathfrak{F}[\mathsf{val}_i, \mathcal{L}]$.

The fact that these conditions are equivalent to each other follows from commutativity of filter-inducing:

$$\begin{aligned}\textstyle\bigcap_{i\in I}\mathfrak{F}[\mathsf{val}_i, \mathcal{L}] = \mathfrak{F}\left[\bigcap_{i\in I}\mathfrak{F}[\mathsf{val}_i], \mathcal{L}\right] &= \mathfrak{F}[\mathfrak{F}[\mathsf{val}], \mathcal{L}] \\ &= \mathfrak{F}[\mathsf{val}, \mathcal{L}].\end{aligned}$$

For the third claim, again $\mathsf{val}(\mathcal{L} \implies \mathcal{K}) = \mathsf{true}$ if and only if $\mathcal{K} \subseteq \mathfrak{F}[\mathsf{val}, \mathcal{L}]$, which means that $\mathfrak{F}[\mathsf{val}, \mathcal{K}] \subseteq \mathfrak{F}[\mathsf{val}, \mathcal{L}]$. Also, $\mathsf{val}(\mathcal{K} \implies \mathcal{M}) = \mathsf{true}$ is equivalent to $\mathsf{val}_{\mathcal{K}}(\mathcal{M}) = \mathsf{true}$; by the first claim, these two facts imply that $\mathsf{val}_{\mathcal{L}}(\mathcal{M}) = \mathsf{true}$ as well. □

We may observe that the third claim in this proposition more or less establishes the "soundness" of meta-implications as a primitive system of logic. In particular, since $\top \implies \mathcal{L}$ is equivalent to $\mathcal{L}$ (under all valuations), then if both $\mathsf{val}(\mathcal{L}) = \mathsf{true}$ and $\mathsf{val}(\mathcal{L} \implies \mathcal{K}) = \mathsf{true}$, we must also have $\mathsf{val}(\mathcal{K}) = \mathsf{true}$.

4.2.5 Convergences and Separations

Ultimately, meta-events and valuations will give way to the topological concepts of convergence and separation, and we now discuss how this will take place. In a non-abstract sense, the kind of convergence we are dealing with is convergence in high-probability of an asymptotic sequence $\mathbf{x}_\eta$ of random elements in a topological space X, in the limit as $n \to \infty$.

In the abstract, though, this is no good, because there are no longer any asymptotic sequences, there is just the Boolean algebra $\mathcal{B}$. This problem is easily resolved: for lack of

a better term, we may define an *abstract element* of any set X to be an algebraic entity, written $\mathbf{x}$, such that for each subset $X_0 \subseteq X$ there exists an event "$\mathbf{x} \in X_0$" in $\mathcal{B}$, and such that the mapping $X_0 \mapsto$ "$\mathbf{x} \in X_0$" is a homomorphism of Boolean algebras from the power set 2^X to $\mathcal{B}$.

As noted, we do not actually need to do this, because we already have a well-defined notion of convergence, but the idea of an abstract element emphasizes an understanding of convergence as a logical concept, rather than something to do with the limit of a sequence. In this view, convergences and separations are special kinds of meta-events:

- $\mathbf{x} \to x$ denotes the meta-conjunction $\{\mathbf{x} \in B\}_{B \ni x}$;
- $\mathbf{x} \nrightarrow x$ denotes the meta-negation (or meta-disjunction) $\mathbf{x} \to x \implies \bot$.

As usual, the $B \ni x$ refers to the set of open neighborhoods of x, and since this set is clearly filtered, then the separation $\mathbf{x} \nrightarrow x$ is easily seen to be equivalent to its previous definition

$$\mathsf{val}(\mathbf{x} \nrightarrow x) = \exists_{B \ni x} \mathsf{val}(\mathbf{x} \notin B).$$

Similarly, the meta-implication

$$\mathbf{x} \to x \implies \mathbf{y} \to y, \tag{4.11}$$

is valuated by

$$\mathsf{val}(\mathbf{x} \to x \implies \mathbf{y} \to y) = \forall_{B_y \ni y} \exists_{B_x \ni x} \mathsf{val}(\mathbf{x} \in B_x \implies \mathbf{y} \in B_y). \tag{4.12}$$

Note that (4.12) closely resembles the definition of a continuous function; in fact, if $\mathbf{y}$ is the image $\mathbf{y} = f(\mathbf{x})$ of a function $f : X \to Y$ which is continuous at x, then the implication (4.11) must hold by definition under any valuation.[12]

The ability to construct meta-implications in this way, and more generally, to temporarily "assume" that $\mathbf{x} \to x$ and work under the relative valuation $\mathsf{val}_{\mathbf{x} \to x}$ will allow us to

[12]Technically, if we really wish to work with "abstract elements" rather than asymptotic sequences, we would have to define what $f(\mathbf{x})$ means; this is not difficult but also not necessary, since the only way a relationship $\mathbf{y} = f(\mathbf{x})$ will ever arise in practice is if two random elements are literally related in this way.

use various common techniques to detect limit points which do not immediately translate to the the less abstract setting of convergence in probability. Indeed, for a non-random sequence x_η, the meta-implication $x_\eta \to x \implies y_\eta \to y$ corresponds to the condition that, for any subsequence along which $x_{\eta'} \to x$, we also have $y_{\eta'} \to y$.

It is not clear whether the assumption $\mathbf{x} \to x$ has an intuitive interpretation with respect to high-probability convergence. But, the fact that all of this is well-defined means that the intuitive notion of a "w.h.p. subsequence" (or lack thereof) is irrelevant, and we may simply think of convergence as a logical condition rather than a limit.

4.2.6 Convergence to Subsets and Compactness

We will make extensive use of the topological property of compactness, and we begin by reviewing some definitions. For a subset $X_0 \in X$:

- an *open cover* of X_0 is a collection of open sets $\{B_i\}_{i \in I}$ such that X_0 is contained in the union $\bigcup_{i \in I} B_i$;
- X_0 is *compact* if every open cover of X_0 contains a *finite sub-cover*, that is, a finite subcollection $\{B_1, \ldots, B_j\}$ which is also an open cover of X_0.

Compactness is a fundamental property with far-reaching consequences; for example, in metric spaces, compactness is equivalent to sequential compactness, meaning that every infinite sequence has a convergent sub-sequence.

In order to exploit compactness, we first extend the definition of convergence to allow convergence to a subset in the natural way:

- $\mathbf{x} \to X_0$ denotes the meta-conjunction $\{\mathbf{x} \in B\}_{B \ni X_0}$,

where B ranges over all open neighborhoods of X_0. While this definition is technically valid for any $X_0 \subseteq X$, makes most sense if X_0 is a closed set. In most cases, we will want the set X_0 to be compact as well, since the ability to choose finite sub-covers somewhat makes up for the inability to form infinite conjunctions.

The following proposition essentially demonstrates that, if X_0 is compact, then the assumption $\mathbf{x} \to X_0$ is "the same" as the assumption that $\mathbf{x} \to x$ for some fixed but unknown element $x \in X_0$. Since there is nothing special about the w.h.p. valuation, we state this proposition in general terms.

Theorem 4.2.4. *For any compact set X_0, any valuation* val*, and any $L \in \mathcal{B}$:*

$$\mathsf{val}(\mathbf{x} \to X_0 \implies L) = \forall_{x \in X_0} \mathsf{val}(\mathbf{x} \to x \implies L). \tag{4.13}$$

In addition:

1. *the same holds of L is replaced by any meta-event $\mathcal{M}$;*
2. *for any subset $X_1 \subseteq X_0$, if $\mathsf{val}(\mathbf{x} \nrightarrow x) = \mathsf{true}$ for all $x \notin X_1$, then $\mathsf{val}(\mathbf{x} \to X_1) = \mathsf{true}$.*

Proof. For any $x \in X_0$, clearly

$$\mathsf{val}(\mathbf{x} \to x \implies \mathbf{x} \to X_0) = \mathsf{true}$$

under any valuation, since every neighborhood of X_0 is also a neighborhood of x. It follows that, if $\mathsf{val}(\mathbf{x} \to X_0 \implies L) = \mathsf{true}$, we also have $\mathsf{val}(\mathbf{x} \to x \implies L) = \mathsf{true}$ for each $x \in X_0$.

The other direction requires the compactness property. Assume the right-hand side of (4.13) valuates to true, in which case, for each $x \in X$, we have a neighborhood $B_x \ni x$ such that

$$\mathsf{val}(\mathbf{x} \in B_x \implies L) = \mathsf{true}.$$

Compactness lets us choose a finite subcover $\{B_i\}$, and we may now form the finite conjunction

$$\begin{aligned} \textstyle\bigwedge_i (\mathbf{x} \in B_i \implies L) &= \textstyle\bigwedge_i (\mathbf{x} \in B_i) \implies L \\ &= (\mathbf{x} \in B) \implies L, \end{aligned}$$

where $B = \bigcup_i B_i$ is an open neighborhood of X_0. It follows that $\mathsf{val}(\mathbf{x} \to X_0 \implies L) = \mathsf{true}$, and the main claim has be proved.

The first additional claim is immediate by proposition, 4.2.3. For the second, we let $\mathcal{M}$ denote the meta-event $\mathbf{x} \to X_1$; in this case we clearly have $\mathsf{val}(\mathbf{x} \to x \implies \mathcal{M}) = \mathsf{true}$ for $x \in X_1$, and for $x \notin X_1$, we have

$$\mathsf{val}(\mathbf{x} \to x \implies \bot \implies \mathcal{M}) = \mathsf{true}.$$

Hence, $\mathsf{val}(\mathbf{x} \to x \implies \mathcal{M})$ for all $x \in X_0$, and the proof is complete. □

At a less abstract level, this theorem has two important consequences (provided, of course, that we are able to show that $\mathsf{val}(\mathbf{x} \to X_0) = \mathsf{true}$ for a compact set X_0). First, in order to show that some meta-event $\mathcal{M}$ occurs, we may perform a sort of case-based reasoning, where we show that the assumption $\mathbf{x} \to x$ implies $\mathcal{M}$ for every individual $x \in X_0$.

Second, by choosing a singleton subset $\{x_1\} = X_1 \subseteq X_0$, the second claim of this theorem allows us to invoke the "Sherlock Holmes" variation of proof by contradiction: "if when you have eliminated the impossible, whatever remains, however improbable, must be the truth." Although this may seem to be a roundabout way of proving that $\mathbf{x} \to x_1$, it is typically easier to establish separation rather than convergence. This is because, in order to prove that $\mathsf{val}(\mathbf{x} \nrightarrow x) = \mathsf{true}$ it suffices to find a *single* fixed neighborhood $B \ni x$ such that $\mathsf{val}(\mathbf{x} \notin B) = \mathsf{true}$. On the other hand, to prove directly that $\mathsf{val}(\mathbf{x} \to x)$ we must prove that $\mathsf{val}(\mathbf{x} \in B_1) = \mathsf{true}$ for *every* fixed neighborhood $B \ni x_1$.

In applications, where the valuation is one of the high-probability conditions, both of this techniques will prove quite valuable when manipulating asymptotic random elements in a topological space. The first allows us to perform "w.h.p. case based reasoning," by temporarily assuming that $\mathbf{x} \to x$ w.h.p., an assumption that does not have an obvious intuitive interpretation. The second allows us to reason by "w.h.p. contradiction" for the purposes of establishing a w.h.p. convergence.

4.2.7 Separation and Hausdorff Spaces

Recall that two subsets X_1, X_2 of a topological space are *separable by neighborhoods* if there exist disjoint open neighborhoods $B_1 \supseteq X_1$ and $B_2 \supseteq X_2$. There are various separation axioms which describe what kinds of subsets in X can be separated by neighborhoods, but the only one we will use is the following:

- X is *Hausdorff* if every pair of distinct points can be separated by neighborhoods.

In the notation of "w.h.p. logic", the fact that x_1, x_2 can be separated by neighborhoods is equivalent to the fact that

$$\mathsf{val}\Big((\mathbf{x} \to x_1, \mathbf{x} \to x_2) \implies \bot\Big) = \mathsf{true}$$

under any valuation. The above expression is by definition equivalent to

$$\mathsf{val}(\mathbf{x} \to x_1 \implies \mathbf{x} \nrightarrow x_2) = \mathsf{true}.$$

Intuitively, then, the Hausdorff assumption allows us to reverse the "Sherlock Holmes" argument. Once we have found the "truth," in the form of a point $x_1 \in X$ such that $\mathsf{val}(\mathbf{x} \to x_1) = \mathsf{true}$, then we can rule out "the impossible," by concluding that $\mathsf{val}(\mathbf{x} \nrightarrow x_2) = \mathsf{true}$ for any $x_2 \neq x_1$. Provided that our valuation is proper, meaning that $\mathsf{val}(\bot) = \mathsf{false}$, this in turn implies that $\mathsf{val}(\mathbf{x} \nrightarrow x_2) = \mathsf{false}$.

As such, the Hausdorff property is also important, though perhaps not quite as vital as compactness. Moreover, most "natural" topological spaces are Hausdorff, but this is not the case for compactness, so ultimately compactness will play a more important role in our techniques than the Hausdorff property.

4.2.8 Logic and "Plain English"

In applications, we will not discuss valuations, filters, and other lattice-theoretic concepts, and instead use the same "plain English" vernacular in which mathematical discourse usually takes place. The translation between the two is fairly clear; for example, both

- if $\mathcal{L}$; then $\mathcal{K}$ w.h.p.
- the following holds w.h.p.: if $\mathcal{L}$; then $\mathcal{K}$

are equivalent to the condition that $\mathsf{whp}(\mathcal{L} \implies \mathcal{K}) = \mathsf{true}$. Moreover, any potentially ambiguous situation can be resolved easily by constructing an exact expression using valuations if necessary. Nevertheless, we briefly discuss some of these issues below.

The Meta-Meta-Level

In effect, there are two layers of "meta-language" involved. The "non-meta" language consists of elements of the Boolean algebra $\mathcal{B}$, and "truth" is the maximal element $\top$. The meta-level is w.h.p. logic, where the logical entities are meta-events and "truth" is the expression true. "Plain English" is the meta-meta-level, where the "truth" is just the "actual truth."

In general, our convention is that each written instance of the term "w.h.p." refers to a unique "w.h.p. logical context," for instance:

- "assume $\mathcal{L}$ w.h.p.; then $\mathcal{K}$ w.h.p." means that if $\mathsf{whp}(\mathcal{L}) = \mathsf{true}$ then $\mathsf{whp}(\mathcal{K}) = \mathsf{true}$;
- "assume $\mathcal{L}$; then $\mathcal{K}$ w.h.p." means that $\mathsf{whp}(\mathcal{L} \implies \mathcal{K})$.

A simple case where these differ is the following. Let $K, L \in \mathcal{B}$, and recall that, although we are using abbreviated notation, each of these is in fact an asymptotic sequence $K_\eta = (K_1, K_2, \ldots)$ of events. Now, assume that K_η and L_η occur independently with probability $1/2$ for each η. In this case, since $\mathsf{whp}(L) = \mathsf{false}$, then the first of the above statements is true. But since $\mathsf{P}[L_\eta \implies K_\eta] = 3/4$ for all η, then we have $\mathsf{whp}(L \implies K) = \mathsf{false}$.

Typically, things outside the w.h.p. context will relate to the problem definition itself. A typical example is:

- assume $\mathbf{G}_\eta$ is an asymptotic sequence of random graphs; then the following holds w.h.p.: ...

In this case, the fact that each $\mathbf{G}_\eta$ is a random element in the set $\mathbb{G}$ of graphs (that is, configurations) is not something that holds "w.h.p." since this would make no sense.

On the other hand, the actual asymptotic parametrization, using the residual distribution, can generally be moved inside the w.h.p. context. As shown in the previous chapter, due to the fact that our convergence assumptions on the residual distribution are based on a metrizable topology, the assumption that $\mu_{\mathbf{G}_\eta} \to \mu$ deterministically can be replaced with convergence in high probability. Hence, an assertion along the lines of "$\boldsymbol{\mathcal{G}}(\mu)$ satisfies a property α w.h.p." is more or less equivalent to $\mathsf{whp}(\mu_{\mathbf{G}_\eta} \to \mu \implies \alpha(\mathbf{G}_\eta)) = \mathsf{true}$.

There are certain technicalities involved, since the convergence $\mu_{\mathbf{G}_\eta} \to \mu$ does not guarantee that the number of vertices grows asymptotically with η, and of course we must assume that $\mathbf{G}_\eta$ is $\mathbb{H}$-conditionally uniform, For these reasons, we will not expand on the discussion of this topic in the previous chapter.

Conditional Probabilities

A w.h.p. implication $L \implies K$ is purely a logical construction, and does not depend on the conditional probability $\mathsf{P}[L_\eta \mid K_\eta]$ as $\eta \to \infty$ Indeed, if $\neg K$ occurs w.h.p., then the implication $K \implies L$ must hold w.h.p. as well, even if $\mathsf{P}[L \mid K] = 0$.

On the other hand, we may compute

$$\begin{aligned}\mathsf{P}[K_\eta \implies L_\eta] = 1 - \mathsf{P}[K_\eta \wedge \neg L_\eta] &= 1 - \mathsf{P}[K_\eta]\mathsf{P}[\neg L_\eta \mid K_\eta] \\ &\geq 1 - \mathsf{P}[\neg L_\eta \mid K_\eta] \\ &= \mathsf{P}[L_\eta \mid K_\eta],\end{aligned}$$

and thus if the conditional probability $\mathsf{P}[L_\eta \mid K_\eta]$ is appropriately high, then the implication $K \implies L$ holds with high probability as well.

It is possible and entirely straightforward to define a "conditional w.h.p. valuation" along these lines, but we do not need this construction. The important points are that $L \implies K$ can occur w.h.p. even though the conditional probability is not high, but if the conditional probability is high than a w.h.p. implication must also hold.

4.3 Topological Representation

We now discuss our second layer of abstraction, in which information is organized topologically rather than as a collection (that is, a filter) of events which occur w.h.p. This involves a *topological representation*, which is an asymptotic sequence of random elements $\mathbf{x}_\eta$ in a topological space X, such that the "relevant" questions can be answered by a w.h.p. convergence $\mathbf{x}_\eta \to x$. As usual, we will drop the subscript "η," and simply call $\mathbf{x}$ a "random element," with the usual conventions and caveats in place.

To a certain extent, the move from logic (or lattices) to topology is a change in nomenclature (i.e. a meta-conjunction is now a convergence). Moreover, since all that matters about the topological representation is if and where it converges, then ultimately the topological nomenclature also becomes unnecessary. We now are searching for a "special" element in the representation space X, and we shall use "constraint satisfaction" terminology to describe this new problem:

- a *w.h.p. solution space* is a compact subset $X_0 \subseteq X$ such that $\mathbf{x} \to X_0$ w.h.p.;
- a *w.h.p. strong solution* is a point $x \in X$ such that that $\mathbf{x} \to x$ w.h.p.;
- a *w.h.p. non-solution* is a point $x \in X$ such that that $\mathbf{x} \nrightarrow x$ w.h.p.;
- a *w.h.p. weak solution* is a point $x \in X$ which is not a w.h.p. non-solution.
- a *w.h.p. solution constraint* is a condition on X which must be satisfied by every w.h.p. weak solution.

In accordance with the "constraint satisfaction" metaphor, we will typically refer to a weak solution as simply a "solution" as in "every solution satisfies...", while a strong solution is a "unique solution." Provided that the compact-Hausdorff assumption discussed below holds, this will be accurate, and the idea is to narrow down possible solutions by imposing constraints until only one possibility remains.

The translation from topological to "constraint satisfaction" language for the first four of the concepts introduced above is immediate by definition. And, using the meta-implications defined in the previous section, the translation of facts about $\mathbf{x}$ to solution constraints is fairly natural as well.

Specifically, for random elements $\mathbf{x}, \mathbf{y} \in X, Y$, a subset $X_0 \subseteq X$, and a function $f : X \to Y$:

- $\mathbf{y}$ is *w.h.p. determined* by $\mathbf{x}$ and f on X_0 if $\mathbf{x} \to x \implies \mathbf{y} \to f(x)$ w.h.p. for each $x \in X_0$;
- $\mathbf{y}$ is *w.h.p. completely determined* by $\mathbf{x}$ if, in addition, X_0 is a w.h.p. solution space.

These conditions are not literally solution constraints, and it is not in general the case that the meta-implication $\mathbf{x} \to x \implies \mathbf{y} \to f(x)$ is equivalent to the "meta-meta-level" assertion that every solution to $(\mathbf{x}, \mathbf{y})$ must satisfy $f(x) = y$. In order for this, and other aspects of our strategy to run smoothly, some technical assumptions are required.

Definition 4.3.1. We say the *compact-Hausdorff assumption* is satisfied if the following hold:

1. there exists some w.h.p. compact solution space X_0;
2. the representation space X is Hausdorff.

In §4.3.1, we explore the consequences of this assumption, but they are what one would expect: there is at least one weak solution, at most one strong solution, a unique weak solution is strong, etc. Moreover, these assumptions are maintained under product spaces which allows us to adopt a "kitchen-sink" approach to topological representation: toss in parameters more or less without discretion, and work everything out in the solution space. In this case, due again to compactness, it is not necessary to "solve" the entire topological representation, as long as we show that all weak solutions imply the same result about the graph property we are ultimately interested in.

Accordingly, while we are treating the topological representation as a single "point" $\mathbf{x} \in X$, in applications, the representation space will be a "big" product space. This product space will include the scaled sample path of a discrete random process, as in the example in §4.1.2. However, as we shall see in the next section, a discrete random process may contain more information than is available from this particular representation. Since the compact-Hausdorff assumption is maintained under products, it is not important to decide exactly what belongs in the topological representation at the "beginning," and we may add or remove parameters "on the fly" as necessary.

4.3.1 The Compact-Hausdorff Assumption

The compact-Hausdorff assumption serves the same purposes which it ordinarily does; we will not belabor this point, and instead list its consequences in somewhat of a "laundry-list" form.

Briefly, compactness is the more important of the two, since this is what allows us to invoke theorem 4.2.4. Also, in most cases, it is not necessary to actually know the identity of the compact solution space X_0; just knowing that one exists will generally suffice.

The Hausdorff assumption, which is usually easy to satisfy, simply insures that, for example, we cannot have more than one strong solution. For technical reasons, it is not sufficient for just the solution space to be Hausdorff, but it is enough if a neighborhood $B \supseteq X_0$ is Hausdorff, since in this case $\mathbf{x} \in B$ holds w.h.p., and we may effectively ignore the rest of the space.

The consequences of the compact-Hausdorff assumption are listed below.[13] As usual, these consequences hold for any (proper) valuation, not just the high-probability valuations, meaning that if we "assume $\mathbf{x} \to x$," this theorem is applicable to the relative w.h.p. valuation. Also, since the compact-Hausdorff assumption will be met in applications, we have made no effort to find the weakest conditions under which any of the following claims

[13] Perhaps none of these claims require proof, since they are basic topological facts. However, in light of our somewhat non-standard definitions, we perform the requisite due diligence.

will hold.

Theorem 4.3.1. *For Hausdorff spaces X, Y, the following hold for any high-probability guarantee or proper valuation.*

1. *For any random element $\mathbf{x} \in X$:*
   (a) *the set of weak solutions is closed;*
   (b) *a strong solution is a unique weak solution;*
   (c) *if $\mathbf{y}$ is determined by $\mathbf{x}$ and f on a subset X_1, then:*
      - *if $x \in X_1$ is a weak solution to $\mathbf{x}$, then (x, y) is a weak solution to $(\mathbf{x}, \mathbf{y})$ if and only if $y = f(x)$;*
      - *the restriction of f the the set of all weak solutions to $\mathbf{x}$ is continuous.*
2. *If $\mathbf{x}$ has compact solution space X_0:*
   (a) *the set of weak solutions is non-empty subset of X_0, and is therefore the minimal compact solution space;*
   (b) *a unique weak solution is a strong solution;*
   (c) *if $\mathbf{y}$ is completely determined by $\mathbf{x}$ and f then $\mathbf{y}$ has compact solution space in some subset of $f(X_0)$.*

Proof (Part 1). (1a) By definition a non-solution satisfies $\mathbf{x} \notin B$ w.h.p. for some $B \ni x$, and every $x' \in B$ is also a non-solution, so the set of non-solutions is open, and its complement, the set of weak solutions, is closed. (1b) If both $\mathbf{x} \to x$ and $\mathbf{x} \nrightarrow x$ w.h.p., then we would have $\perp$ w.h.p., which is not possible. Conversely, the Hausdorff assumption implies that

$$(\mathbf{x} \to x_1, \mathbf{x} \to x_2) \implies \perp,$$

for $x_1 \neq x_2$, so if $\mathbf{x} \to x_1$ w.h.p. then x_2 is a non-solution by definition.

(1c) For the first claim, if $(x, f(x))$ is not a weak solution, then we have

$$\mathbf{x} \to x \implies (\mathbf{x}, \mathbf{y}) \to (x, f(x)) \implies \perp$$

which means x is not a weak solution. Uniqueness follows similarly to the above from the Hausdorff property on Y.

For the second claim, assume w.l.o.g. that every $x \in X_1$ is a weak solution (but there may be others), let $Z \subseteq X \times Y$ denote the (closed) set of weak solutions to $(\mathbf{x}, \mathbf{y})$, and for any $Y_1 \subseteq Y$, let

$$X_{Y_1} = \{x \in X : (x, y) \in Z \text{ for some } y \in B\}.$$

In particular, if Y_1 is closed, then so is $Z \cap X \times Y_1$, and therefore X_{Y_1}, which is the projection onto the first coordinate, is also closed. Now, if $x \in X_1$, then $(x, f(x))$ is the unique element of Z with first coordinate equal to x. If follows $x \in X_{Y_1} \cap X_1$ if and only if $f(x) \in Y_1$.

Since this is true for any closed subset $Y_1 \subseteq Y$, then for any open set $B = Y \backslash Y_1$, if $f(x) \in B$ then $x \in X_1 \backslash X_{Y_1}$, which is an open subset of X_1; thus, f is continuous when restricted to X_1. □

Proof (Part 2). First, by theorem 4.2.4, the fact that X_0 is a compact solution space means that, for any meta-event $\mathcal{M}$, if $\mathbf{x} \to x \implies \mathcal{M}$ w.h.p. for every fixed $x \in X_0$, then $\mathcal{M}$ holds w.h.p. (2a) Choose the meta-event $\mathcal{M} = \perp$, in which case, if there are no weak solutions $\perp$ would occur w.h.p. Next, for any $x_1 \notin X_0$, we choose the meta-event $\mathcal{M} = \mathbf{x} \nrightarrow x_1$. By the Hausdorff condition, $\mathbf{x} \to x \implies \mathcal{M}$ for any $x_1 \neq x \in X_0$, and hence x_1 is a non-solution. It follows that the set of weak solutions is a non-empty subset of X_0, and since this set is closed, it is compact. And, since this holds for any compact solution space X_0, it follows that the set of weak solutions is the minimal compact solution space.

(2b) If x is a unique weak solution, then $\{x\}$ is a compact solution space, so x is strong solution. (2c) We may assume w.l.o.g. that X_0 is minimal, in which case every $x \in X_0$ is a weak solution, and as shown above, f is thus continuous from $X_0 \to Y$. Hence, $f(X_0)$ is a compact solution space for $\mathbf{y}$. □

4.3.2 Structural and Probabilistic Constraints

We may identify two different kinds of solution constraints:

- a *structural constraint* has nothing to do with randomness, and only depends on the topological structure of the representation space;
- a *probabilistic constraint* depends on the actual characteristics of $\mathbf{x}$.

The majority of the constraints we discuss in this section (or this chapter) will be structural; in fact, the only real probabilistic constraint we will make use of is martingale concentration (see section 4.5).

Structural constraints will generally hold, not only w.h.p., but *ultimately always (u.a.)*. For structural constraints, the notation of w.h.p. implications serves primarily as a tool to to describe the topology of the representation space. Structural constraints will typically be used to establish one of the following:

- that the image $f(\mathbf{x})$ of $\mathbf{x}$ under a mapping $f : X \to Y$ is completely determined by $\mathbf{x}$, meaning that $\mathbf{x} \to x \implies f(\mathbf{x}) \to x$ u.a. (or w.h.p.);
- that convergence under one topology $\mathcal{T}_1$ implies convergence under a second topology $\mathcal{T}_1$, meaning that $\mathbf{x} \xrightarrow{\mathcal{T}_1} x \implies \mathbf{x} \xrightarrow{\mathcal{T}_2} x$ u.a. (or w.h.p.).

Moreover, the second of these can be understood as a special case of the first, by letting f denote the identity mapping from $(X, \mathcal{T}_1) \to (X, \mathcal{T}_2)$.

Now, a function $f : X \to Y$ is continuous, then $\mathbf{x} \to x \implies f(\mathbf{x}) \to x$ u.a. by definition. Moreover, by Theorem 4.3.1, if $f(\mathbf{x})$ is completely determined by $\mathbf{x}$, then it must be the case that f is continuous, when restricted to the minimal compact solution space.

However, $f(\mathbf{x})$ may be completely determined by $\mathbf{x}$ even if $f : X \to Y$ is not continuous on all of X. Hence, the kind of situation that requires non-trivial structural constraints is where f is continuous when restricted to some compact solution space $X_0 \subseteq X$, but not on the entire representation space X. In this case, w.h.p. determination does not follow immediately from continuity, and must be established by other means.

4.3.3 Convergence Metric Spaces

We now discuss the special case where the domain X is a metric space $X = (X, \rho)$. Metric spaces possess certain useful characteristics, for example, every metric space is not only Hausdorff, but also *normal*, meaning that any disjoint closed sets can be separated by neighborhoods. Most of the spaces we work with will be either metric or at least metrizable.

Convergence in a metric space can be in terms of the metric, since $\mathbf{x} \to x$ is equivalent to $\rho(\mathbf{x}, x) \to 0$ by definition. In many cases, the metric notion of convergence is more versatile than the pure topological convergence (though perhaps less elegant). For example, certain concepts such as the rate of convergence, or the uniformity of multiple simultaneous convergences, are not well defined in a general topological sense.

Another advantage of metric spaces is that, while topological convergence must generally occur to a fixed limit $x \in X$, in a metric space we may compare two random elements to each other by examining the behavior of $\rho(\mathbf{x}_1, \mathbf{x}_2)$. In particular, if $\rho(\mathbf{x}_1, \mathbf{x}_2) \to 0$, then these random elements determine each other, and it is easily verified that the converse is true as well.

Proposition 4.3.2. *Let $\mathbf{x}_1, \mathbf{x}_2$ be random elements in a metric space X, and assume $\mathbf{x}_1 \to X_0$ w.h.p. for some compact X_0. Then the following are equivalent:*

1. *$\rho(\mathbf{x}_1, \mathbf{x}_2) \to 0$ w.h.p.;*
2. *$\mathbf{x}_2$ is completely determined w.h.p. $\mathbf{x}_1$ (and the identity function).*

Proof. The forward implication is trivial; for the reverse implication, for any fixed $x \in X_0$, we have

$$\mathbf{x}_1 \to x \implies \mathbf{x}_2 \to x \implies \rho(\mathbf{x}_1, x) + \rho(\mathbf{x}_2, x) \to 0,$$

and thus $\rho(\mathbf{x}_1, \mathbf{x}_2) \to 0$ by the triangle inequality. □

4.3.4 Equicontinuity and Uniform Convergence

We now consider spaces of functions from the compact interval $[0,1]$ to a metric (or topological) space X. Our notation is as follows:

- $\mathcal{M}([0,1],X)$ is the space of functions $[0,1] \to X$;
- $\mathcal{C}([0,1],X)$ is the space of functions $[0,1] \to X$;
- elements of either are written as "tuples" $(x_\xi) = (x_\xi)_{\xi\in[0,1]}$;
- the product topology is the default, and uniform convergence is expressed by $(\mathbf{x}_\xi) \xrightarrow{\infty} (x_\xi)$.

These spaces will be used primarily to represent the scaled sample path of a discrete random process, as described in the next section.

Evidently, the choice of the interval $[0,1]$ is without loss of generality, and we may just as well choose any other interval. Also, we will use the same general definitions and conventions for intervals which are not compact; for instance, the product topology is also the default on $\mathcal{M}((0,1),X)$, and elements in this set are written as tuples. Clearly, though, the Arzelà-Ascoli theorem proved below, or any other results which assume a compact interval will no longer be valid.

The product topology has the advantage of preserving compactness, but, as mentioned during the example in §4.1.2, the space $\mathcal{M}([0,1],X)$ under the product topology is not really a "function space." The space $\mathcal{C}([0,1],X)$ of continuous functions $[0,1] \to X$, under the product topology, is not much better; one problem this spaces suffers from is that it is not closed.

The "correct" topology here is the uniform topology, and switching to the uniform topology of course means that compactness needs to be reestablished. In this setting, compactness with respect to the uniform topology is more or less the same as equivalence between the product and uniform topologies, and the basic way to establish this is the Arzelà-Ascoli Theorem.

The "textbook" version of this theorem does not quite meet our literal requirements in this case, so we shall reproduce the proof.[14] The key property is equicontinuity, but since our method of time-scaling (see the next section or the example in §4.1.2) creates jump discontinuities, we use the following condition instead.

Definition 4.3.2. A scaled process $(\mathbf{x}_\xi)$ is *asymptotically equicontinuous* if, for any fixed $\xi \in [0,1]$ and $\epsilon > 0$ there exists a neighborhood $B \ni \xi$ (in $[0,1]$) such that

$$\sup\{\rho(\mathbf{x}_\xi, \mathbf{x}_\zeta) : \zeta \in B\} < \epsilon \quad \text{u.a.} \tag{4.14}$$

The "u.a." in this definition may of course be replaced with any other high-probability guarantee, but the default condition here is u.a. This is because the asymptotic equicontinuity typically depends on the specification of a process, and are best understood as a structural assumption. As usual, if the assumption holds with a weaker guarantee, then so do the conclusions we may draw, but for simplicity, we stick with the u.a. condition.

One evident consequence of equicontinuity is that any solution must be a continuous function. A second consequence, as we now show, is that pointwise convergence implies uniform convergence.

Theorem 4.3.3 (Arzelà-Ascoli)**.** *If* $(\mathbf{x}_\xi)$ *is asymptotically equicontinuous then*

$$(\mathbf{x}_\xi) \to (x_\xi) \implies (\mathbf{x}_\xi) \xrightarrow{\infty} (x_\xi) \quad \textit{u.a.}$$

for any (continuous) function (x_ξ)*.*

Proof. Omitting the "u.a." for simplicity, we must show that for all $\epsilon > 0$,

$$(\mathbf{x}_\xi) \to (x_\xi) \implies \sup\{\rho(\mathbf{x}_\xi, x_\xi) : \xi \in [0,1]\} < \epsilon.$$

[14] It would not be difficult to get around the minor technicalities which prevent a "cut-and-dry" application of Arzelà-Ascoli . The issue is that our method of scaling discrete sample paths does not produce continuous functions, due to the use of the least integer function, this could be overcome e.g. by interpolating. Nevertheless, since this is an important part of the general construction, it is just as well that we give a proof.

Using the equicontinuity of $(\mathbf{x}_\xi)$ and the continuity of (x_ξ), for any $\xi \in [0,1]$, we may choose a neighborhood $B_\xi \ni \xi$ sufficiently small that

$$\sup\{\rho(\mathbf{x}_\xi, \mathbf{x}_\zeta) + \rho(x_\xi, x_\zeta) : \zeta \in B_\xi\} < \epsilon/2$$

Also, pointwise convergence $(\mathbf{x}_\xi) \to (x_\xi)$ implies that $\rho(x_\xi, \mathbf{x}_\xi) < \epsilon/2$, and by the triangle inequality, we thus have

$$(\mathbf{x}_\xi) \to (x_\xi) \implies \sup\{\rho(\mathbf{x}_\zeta, x_\zeta) : \zeta \in B_\xi\} < \epsilon$$

for every fixed $\xi \in [0,1]$.

Since the interval $[0,1]$ is compact, we may choose a finite subcover $\{B_i\}$ of the open cover B_ξ (of the interval $[0,1]$), and form the finite conjunction:

$$(\mathbf{x}_\xi) \to (x_\xi) \implies \bigwedge_{i\in I}\Big(\sup\{\rho(\mathbf{x}_\zeta, x_\zeta) : \zeta \in B_i\} < \epsilon\Big) \implies \sup\{\rho(\mathbf{x}_\zeta, x_\zeta) : \zeta \in [0,1]\} < \epsilon.$$

□

4.3.5 Summability and Convergence in ℓ_1

Another space which we will work with frequently is the countable space $\mathbb{R}^\omega$. Our notation is as follows:

- $\mathbb{R}^\omega$ denotes any countable product of $\mathbb{R}$, and we typically use the set $\mathbb{Z}^* = \{0, 1, 2, \ldots\}$ as our countable index set;
- we write elements of $\mathbb{R}^\omega$ in "function" notation, as in $x = (x(0), x(1), \ldots)$;
- $\ell_1 \subseteq \mathbb{R}^\omega$ consists of all elements with $\|x\|_1 < \infty$;
- pointwise convergence is again the default, and convergence in ℓ_1 is expressed by $\mathbf{x} \xrightarrow{\ell_1} x$.

The space $\mathbb{R}^\omega$ is most often used to parametrize the configuration model using the degree and/residual residual distribution, both of which are elements in the set $\Phi = \mathrm{Dist}(\mathbb{Z}^*) \subseteq \ell_1$

of distributions on $\mathbb{Z}^*$. While uniform convergence is obviously well-defined in $\mathbb{R}^\omega$, we will work almost exclusively with either the stronger ℓ_1 topology or the weaker product topology.

The situation is roughly analogous to what we faced above in the space $\mathcal{M}([0,1], X)$: pointwise convergence is the default, but what we really want is ℓ_1 convergence. The property that

$$\mathbf{x} \to x \implies \mathbf{x} \xrightarrow{\ell_1} x$$

is analogous to equicontinuity, and we will call this property *summability*. This property, along with the relationship between the ℓ_1 and product topologies on $\mathbb{R}^\omega$, was discussed in chapter 2, and the situation for asymptotic random elements is basically the same.

Proposition 4.3.4. *Let* $\mathbf{x}$ *be a random element in* $\mathbb{R}^\omega$*, and assume* $\|\mathbf{x}\|_1 < c$ *u.a. for some constant* c*. Then the following are equivalent:*

1. $\mathbf{x} \to x \implies \mathbf{x} \xrightarrow{\ell_1} x$ *u.a. for every* $x \in \ell_1$*,*
2. $\mathbf{x} \to x \implies \|\mathbf{x}\|_1 \to \|x\|_1$ *u.a. for every* $x \in \ell_1$*,*
3. *for every* $\epsilon > 0$*, there exists* J *such that* $\sum_{i>J} |\mathbf{x}(i)| < \epsilon$ *u.a.*

If any of these hold, we say $\mathbf{x}$ *is* summable*.*

Proof. The equivalence of the first two conditions is a consequence of the fact that the ℓ_1 topology is equivalent (on ℓ_1, and not on all of $\mathbb{R}^\omega$) to the topology induced by the product topology and the map $x \mapsto \|x\|_1$, as shown in section 2.2 of chapter 2. And, it is clear that the third claim implies the first due to the standard triangle inequality argument.

For the opposite implication, note that the fact that $\|\mathbf{x}\|_1 < c$ u.a. implies that $|\mathbf{x}(i)| < c$ u.a. for each i, which implies that $\mathbf{x}$ has compact state space u.a. with respect to the product topology. Also, clearly every solution must also belong to ℓ_1, since if $\|x\|_1 = \infty$ then $\mathbf{x} \to x \implies \|\mathbf{x}\|_1 \to \infty$.

Now, for every $\epsilon > 0$ and $x \in \ell_1$, the fact that $\mathbf{x} \to x \implies \|\mathbf{x}\|_1 \to \|x\|_1$ u.a. means that there exists J_x such that $\mathbf{x} \to x \implies \sum_{i>J_x} |\mathbf{x}(i)| < \epsilon$ u.a. This in turn implies

that there exist an neighborhood $B_x \ni x$ (with respect to the product) topology such that $\mathbf{x} \in B_x \implies \sum_{i>J_x} |\mathbf{x}(i)| < \epsilon$. Third condition can now be achieved by choosing a finite subcover of the compact solution space and letting $J = \max\{J_{x_1}, \ldots, J_{x_k}\}$. □

4.3.6 Uniform Summability

We now consider the space $\mathcal{M}([0,1], \mathbb{R}^\omega)$ of mappings from $[0,1] \to \mathbb{R}^\omega$. There are various possible topologies on this space, but we are primarily interested in the following two:

- the *product-product* topology is the default, so $(\mathbf{x}_\xi) \to (x_\xi)$ means that $\mathbf{x}_\xi(i) \to x_\xi(i)$ for every fixed pair $\xi \in [0,1]$ and $i \in \mathbb{Z}^*$;
- the *uniform-ℓ_1* topology is the uniform topology over $\xi \in [0,1]$ with respect to the ℓ_1 topology on $\mathbb{R}^\omega$; convergence in this topology is written $(\mathbf{x}_\xi) \xrightarrow{\infty,\ell_1} (x_\xi)$, which means that

$$\sup\{\|\mathbf{x}_\xi - x_\xi\|_1 : \xi \in [0,1]\} \to 0.$$

We call *uniform summability* the condition that the first above convergence implies the second.

Proposition 4.3.5. *Let $(\mathbf{x}_\xi)$ be a random element in $\mathcal{M}([0,1], \mathbb{R}^\omega)$ and assume that*

1. *$(\mathbf{x}_\xi(i))$ is asymptotically equicontinuous for each $i \in I$;*
2. *for every $\epsilon > 0$, there exists J such that $\sup\{\sum_{i>J} |\mathbf{x}_\xi(i)| : \xi \in [0,1]\} < \epsilon$ u.a.*

Then:

1. *the function $(\mathbf{x}_\xi)$ is asymptotically equicontinuous in ℓ_1;*
2. *for any $(x_\xi) \in \mathcal{C}([0,1], \ell_1)$,*

$$(\mathbf{x}_\xi) \to (x_\xi) \implies (\mathbf{x}_\xi) \xrightarrow{\infty,\ell_1} (x_\xi) \quad u.a.$$

In this case, we say $(\mathbf{x}_\xi)$ *is* uniformly summable.

Proof. Due to coordinatewise equicontinuity, for any fixed J, the first J coordinates of $(\mathbf{x}_\xi)$ clearly constitute an equicontinuous random element in $\mathcal{C}([0,1],\ell_1)$, and the uniform bound on $\sup\{\sum_{i>J}|\mathbf{x}_\xi(i)| : \xi \in [0,1]\}$ thus implies equicontinuity in ℓ_1. The second claim is immediate. □

4.4 Topological Representation of a Random Process

We now discuss the topological representation of random processes, which are the central random structures in the algorithmic analysis of random graphs. The basic idea was presented in §4.1.2: we scale both the state space and the time scale so as to resemble a continuous function. However, a discrete process often contains information which is not "visible" with respect to this representation, and it is often necessary to include some additional parameters in the overall topological representation.

For the purposes of this section, we shall assume that the state space has already been scaled; the random object we are dealing with is thus the sample path $(\mathbf{x}_t) = (\mathbf{x}_0, \ldots, \mathbf{x}_T)$ for which the state space is some topological space X. We briefly review the time-scaling procedure from §4.1.1:

- we define a mapping $\xi \mapsto \mathbf{x}_{\lfloor \xi T \rfloor}$ for $\xi \in [0,1]$, where $\lfloor \xi T \rfloor = \max\{t \in \mathbb{Z} : t \leq \xi T\}$;
- we abbreviate by $\mathbf{x}_{\lfloor \xi T \rfloor} = \mathbf{x}_\xi$, and use "tuple" notation $(\mathbf{x}_\xi)$ to denote this function.

The representation space is thus space $\mathcal{M}([0,1],X)$, and as usual the product topology is the default.

4.4.1 Increments, Integrals, and Derivatives

The *increment* of an $\mathbb{R}$-valued process at time $t \geq 1$ is defined by $\Delta\mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$. For consistency, we may also define $\Delta\mathbf{x}_0 = 0$ and call the process $(\Delta\mathbf{x}_t)$ the *increment process.*

The state at time t can thus be expressed as the sum of increments:

$$\mathbf{x}_t = \mathbf{x}_0 + \sum_{s=1}^{t} \Delta \mathbf{x}_s,$$

When scaled, this expression closely resembles the definition of the Riemann integral, and this similarity is the basis of the differential (or integral) equations.

In fact, we may observe that

$$\mathbf{x}_t = \mathbf{x}_0 = \int_{s=0}^{t} \Delta \mathbf{x}_s$$

holds precisely for any integer t. When a process is time-scaled, we must multiply the increments by T to arrive at the correct formula, which is,

$$\mathbf{x}_\xi = \mathbf{x}_{\lfloor \xi T \rfloor} = \int_{\zeta=0}^{\lfloor \xi T \rfloor} T \cdot \Delta \mathbf{x}_{\lfloor \zeta T \rfloor}$$

where $\Delta \mathbf{x}_{\lfloor \zeta T \rfloor} = \mathbf{x}_{\lfloor \zeta T \rfloor} - \mathbf{x}_{\lfloor \zeta T \rfloor - 1}$.

Now, due to the jump-discontinuities created by our method of time-scaling, the scaled sample path $(\mathbf{x}_\xi)$ is not differentiable, but the "moral" derivative of this function is $T \cdot \Delta \mathbf{x}_{\lfloor \xi m \rfloor}$. This problem could be overcome by, say, using linear interpolation to create a continuous representation of the scaled sample path, but this gives rise to other technicalities, and using step-functions ultimately works just as well.

Bounds on Increment Sizes

In order for this to work out rigorously, the increments of a process must satisfy some basic conditions. Hence, for any fixed constant c, we define:

- $(\mathbf{x}_t)$ is *Lipschitz* if $\sup\{|\Delta \mathbf{x}_t| / T : 1 \leq t \leq T\} < c$ u.a.;
- $(\mathbf{x}_t)$ has *asymptotically bounded variation* if $\sum_{t=1}^{T} |\Delta \mathbf{x}_t| < c$ u.a.;
- $(\mathbf{x}_t)$ is *well-scaled* if it has asymptotically bounded variation, and the time-scaled function $(\mathbf{x}_\xi)$ is asymptotically equicontinuous.

As with asymptotic equicontinuity, the "u.a." may be replaced with any other high-probability guarantee, but again these will generally hold u.a., as they are usually determined by the specification of a process, rather than the outcome.

Any solution to a process satisfying one of the above conditions must behave accordingly:

- a function $(x_\xi) \in \mathcal{M}([0,1],\mathbb{R})$ is *Lipschitz-continuous* if there exists a constant c such that $|x_\xi - x_\zeta| \leq c \cdot |\xi - \zeta|$ for all $\xi, \zeta \in [0,1]$;
- a function $(x_\xi) \in \mathcal{M}([0,1],\mathbb{R})$ has *bounded variation* if it can be expressed as the difference between two non-decreasing functions:

  $$x_\xi = x_\xi^+ - x_\xi^-,$$

  and the *total variation* is the infimum of $(x_1^+ - x_0^+) + (x_1^- - x_0^-)$ for all such decompositions.[15]

Proposition 4.4.1. *For any process* $(\mathbf{x}_t)$*:*

1. *if* $(\mathbf{x}_t)$ *is Lipschitz, and the specified duration satisfies* $T \to \infty$*, then any solution* (x_ξ) *is Lipschitz with respect to the same constant* c*;*
2. *if* $(\mathbf{x}_t)$ *has asymptotically bounded variation, then every solution has bounded variation, and the total variation is bounded by the same constant* c*.*

Proof. The Lipschitz bound is immediate. For the bound on total variation, note that $(\mathbf{x}_t)$ can be decomposed into monotonic components by letting $\mathbf{x}_t^+ = \mathbf{x}_0 - \sum_{s=1}^t \max\{\Delta\mathbf{x}_s, 0\}$, and $\mathbf{x}_t^- = \mathbf{x}_t - \mathbf{x}$. Any solution to these component processes must clearly be monotonic and absolutely bounded accordingly, and we may therefore decompose any solution in the same way. □

[15] This is perhaps not the standard definition of total variation. This would be the infimum of $\sum_i \left|x_{xi_{i+1}} - \xi_i\right|$ over all finite partitions $(\xi_0 = 0, \ldots, \xi_j = 1)$ of the interval $[0,1]$. But, these are equivalent (see, e.g. [60, 10]), and the above suits our purposes.

It is easily verified that the standard arithmetic operations preserve these conditions.

Proposition 4.4.2. *Let* $(\mathbf{x}_t)$ *and* $(\mathbf{y}_t)$ *be Lipschitz or well-scaled with state space* $\mathbf{x}_t, \mathbf{y}_t \in [C_1, C_2]$ *u.a. for some fixed interval* $[C_1, C_2] \subseteq \mathbb{R}$*. Then:*

1. $(\mathbf{x}_t + \mathbf{y}_t)$ *and* $(\mathbf{x}_t \cdot \mathbf{y}_t)$ *both satisfy the same condition (Lipschitz or well-scaledness);*
2. *if* $f : [C_1, C_2] \to \mathbb{R}$ *is Lipschitz-continuous, then* $(f(\mathbf{x}_t))$ *satisfies the same condition;*

Proof. Immediate. □

4.4.2 The Riemann-Stieltjes Integral

The simplest way to exploit the resemblance between increments and derivatives is perhaps to use the Riemann integral. For the purposes of expediency, though, we will move directly to the more general setting of the Riemann-Stieltjes integral. We shall not review the definition of the Riemann-Stieltjes integral, or get into necessary and sufficient conditions for its existence; for this we refer the reader to any number of texts (e.g. [60, 10, 42]).

The basic facts and notations are:

- we denote the Riemann-Stieltjes integral by $(\int x_\xi dy_\xi)$;
- if (x_ξ) is continuous and (y_ξ) has bounded variation then $(\int x_\xi dy_\xi)$ exists (but not only if);
- if $(\int x_\xi dy_\xi)$ exists, then, due to integration by parts, so does $(\int x_\xi dy_\xi) = (x_\xi \cdot y_\xi) - (\int y_\xi dx_\xi)$;
- if (y_ξ) has bounded variation, then the mapping $(x_\xi) \mapsto (\int x_\xi dy_\xi)$ is continuous with respect to the ℓ_∞ norm.

If the function $y_\xi = \xi$ is the identity, then we just have the the Riemann integral, which we denote by $(\int x_\xi) = (\int x_\xi d\xi)$.

The discrete summation corresponding to the integral $(\int x_\xi dy_\xi)$ is

$$\mathbf{z}_t = \sum_{s=1}^{t} \mathbf{x}_s \cdot \Delta \mathbf{y}_s, \tag{4.15}$$

or perhaps $\sum_{s=0}^{t-1} \mathbf{x}_s \Delta \mathbf{y}_{s+1}$; these indexing technicalities will prove to be irrelevant. However, since the time-scaled functions $(\mathbf{x}_\xi)$ and $(\mathbf{y}_\xi)$ share discontinuities, then it is not well-defined to write $(\int \mathbf{x}_\xi d\mathbf{y}_\xi)$, but this is the idea, and we now show that this summation will converge as expected.

Proposition 4.4.3. *Assume $(\mathbf{x}_t)$ is asymptotically equicontinuous, $(\mathbf{y}_t)$ be well-scaled, and let $\mathbf{z}_t = \sum_{s=1}^{t} \mathbf{x}_s \cdot \Delta \mathbf{y}_s$. Then, for any (x_ξ) and (y_ξ):*

$$\Big((\mathbf{x}_\xi) \to (x_\xi), (\mathbf{y}_\xi) \to (y_\xi)\Big) \implies (\mathbf{z}_\xi) \to \big(\textstyle\int x_\xi dy_\xi\big) \quad u.a.$$

In this case, the same holds if we stagger the summation and instead define $\mathbf{z}_t = \sum_{s=0}^{t-1} \mathbf{x}_s \Delta \mathbf{y}_{s+1}$.

Proof. Since $(\mathbf{x}_t)$ and $(\mathbf{y}_t)$ are equicontinuous, then the two convergences on the left hold uniformly by Arzelà-Ascoli . This, along with the total variation bound on $(\mathbf{y}_t)$, allows us to replace $(\mathbf{x}_t)$ with the discrete sequence $(x_{t/T})$ in the expression for $\mathbf{z}_t$, and compute

$$\begin{aligned}\sup_{0\le t\le T}\left|\mathbf{z}_t - \sum_{s=1}^{t} x_{s/T}\Delta \mathbf{y}_s\right| &\le \sum_{s=1}^{T} \big|(\mathbf{x}_s - x_{s/T})\Delta \mathbf{y}_s\big| \\ &\le \left(\sup_{0\le s\le T}\big|\mathbf{x}_s - x_{s/T}\big|\right)\sum_{s=1}^{T}|\Delta \mathbf{y}_s| \to 0.\end{aligned}$$

This summation can be expressed as a Riemann-Stieltjes integral (by definition) as

$$\sum_{s=1}^{t} x_{s/T}\Delta \mathbf{y}_s = \int_{s=0}^{t} x_{s/T} d\mathbf{y}_{\lfloor s\rfloor} = \int_{\zeta=0}^{t/T} x_\zeta d\mathbf{y}_\zeta,$$

and we thus have $\big\|(\mathbf{z}_\xi) - (\int x_\xi d\mathbf{y}_\xi)\big\|_\infty \to 0$.

Hence, it suffices to show that $(\int x_\xi d\mathbf{y}_\xi) \xrightarrow{\infty} 0$. Since we have not imposed a bound on the total variation of $(\mathbf{x}_t)$, then it may be the case that (x_ξ) has unbounded variation.

However, (x_ξ) is indeed continuous and can therefore be approximated uniformly by functions with bounded variation. Hence, for any $\epsilon > 0$, choose (w_ξ) which is continuous with bounded variation such that $\|(x_\xi) - (w_\xi)\|_\infty < \epsilon$, and in this case

$$\left\|(\int x_\xi d\mathbf{y}_\xi) - (\int w_\xi d\mathbf{y}_\xi)\right\|_\infty < \epsilon \sum_{s=1}^{T} |\Delta \mathbf{y}_s| \,.$$

We now integrate by parts:

$$\begin{aligned}(\int w_\xi d\mathbf{y}_\xi) = (w_\xi \cdot \mathbf{y}_\xi) - (\int \mathbf{y}_\xi dw_\xi) &\to (w_\xi \cdot y_\xi) - (\int y_\xi dw_\xi) \\ &= (\int w_\xi dy_\xi).\end{aligned}$$

Since (y_ξ) also must have bounded variation, the proof is complete by making $\epsilon > 0$ arbitrarily small.

To see that the alternate summation $\mathbf{z}_t = \sum_{s=0}^{t-1} \mathbf{x}_s \Delta \mathbf{y}_{s+1}$ satisfies the same conditions, note that, for any fixed $\epsilon > 0$, asymptotic equicontinuity implies that $\sup_{1 \le t \le T}\{|\mathbf{x}_t - \mathbf{x}_{t-1}|\} < \epsilon$ u.a.; this should not be confused with a Lipschitz condition, which requires a much stronger bound of $|\mathbf{x}_t - \mathbf{x}_{t-1}| < c/T$ for a constant c. The weaker bound of $\sup_{1 \le t \le T}\{|\mathbf{x}_t - \mathbf{x}_{t-1}|\} < \epsilon$, though, ensures that

$$\sum_{s=1}^{T} |(\mathbf{x}_s - \mathbf{x}_{s-1})\Delta \mathbf{y}_s| < \epsilon \sum_{s=1}^{T} |\Delta \mathbf{y}_{s+1}|$$

for any $\epsilon > 0$, and hence this alternate summation converges to the same limit. □

The ordinary Riemann integral follows as a corollary; we state this corollary directly in terms of the increment process, which needs to be "unscaled" as follows.

Corollary 4.4.4. *If $(\mathbf{y}_t)$ is asymptotically equicontinuous, then*

$$(T \cdot \Delta \mathbf{x}_{\lfloor \xi T \rfloor}) \to (y_\xi) \implies (\mathbf{x}_\xi) \to (\int y_\xi) \quad u.a.$$

Proof. Immediate. □

4.4.3 Derivatives and Lebesgue Measure

We will often express integral equations in differential form, as in

$$dz_\xi = x_\xi dy_\xi. \tag{4.16}$$

This is primarily for convenience, since it allows us, for example, to express constraints that only hold under certain circumstances more succinctly. It is not immediately obvious what exactly the expression "dz_ξ" signifies, but provided that we do not use this convention abusively, the meaning is clear. The standard interpretation would perhaps be as a so-called "weak-derivative," or else as a Lebesgue measure, either of which involve equivalence modulo a set of measure zero; due to the following theorem, this interpretation is sound.

Theorem 4.4.5 (Lebesgue)**.** *Any function with bounded variation is differentiable outside a set of Lebesgue measure* 0.[16]

Proof. See, e.g., [10]. □

However, the derivative serves another important role, which is to describe the behavior of a function in a small neighborhood of ξ. Since we are using a function (x_ξ) as a representation of a random process, then the time coordinate ξ has a natural direction, and we are specifically interested in the behavior in the immediate future, which means $\zeta \in (\xi, \xi + \epsilon)$. We thus restrict the possible versions of the derivative as follows.

Definition 4.4.1. We shall call (dx_ξ) the *right-differential* of (x_ξ), and define this to be the equivalence class of functions mapping $[0, 1) \to [-\infty, \infty]$ such that:

$$dx_\xi \leq \limsup_{\epsilon \to 0^+} \frac{x_{\xi+\epsilon} - x_\xi}{\epsilon}$$
$$dx_\xi \geq \liminf_{\epsilon \to 0^+} \frac{x_{\xi+\epsilon} - x_\xi}{\epsilon}.$$

[16] Technically, or perhaps traditionally, Lebesgue's theorem states that all monotonic functions are almost-everywhere differentiable, and some additional work is required to generalize to functions of bounded variation. However, due to our definition of bounded variation, this generalization is immediate.

The point here is not to give the most robust definition possible, and attempt to push it to a breaking point. All we hope to accomplish is, for example, to conclude from an inequality $dx_\xi > 0$ that the function (x_ξ) is "right-increasing" at the point ξ. In a more abstract sense, we will describe the "immediate future" as follows.

Definition 4.4.2. For any function (x_ξ) from $[0,1]$ to any set X, the *right-continuation* at a point $\xi \in [0,1)$, which we denote by x_{ξ^+}, is the equivalence class of functions modulo the relation

$$x_{\xi^+} = y_{\xi^+}$$

of there exists $\epsilon > 0$ such that $x_\zeta = y_\zeta$ for all $\zeta \in (\xi, \xi + \epsilon)$.

Again, we will not elaborate, because the use of these definitions will be obvious. For instance $x_{\xi^+} = 0$ means that $x_\zeta = 0$ for $\zeta \in (\xi, \xi + \epsilon)$, while $x_{\xi^+} > 0$ means $x_\zeta > 0$ in such an interval, and so on. Some basic facts and examples:

- dx_ξ respects right-hand limits, meaning that $dx_\xi = \lim_{\zeta \to \xi^+} dx_\zeta$ if this limit exists;
- if (x_ξ) is non-negative and $x_\xi = 0$ then $dx_\xi \geq 0$;
- if $dx_\xi > 0$ then $x_{\xi^+} > x_\xi$;
- if $x_{\xi^+} = x_\xi$ then $dx_\xi = 0$;
- the continuation is well-defined for functions in any set X (not just topological spaces);
- if (x_ξ) and (y_ξ) are continuous functions from $[0,1]$ to a Hausdorff space X, then then the following are equivalent:
  - $y_\xi = x_\xi$ for all $\xi \in [0,1]$,
  - $x_0 = y_0$, and $x_{\xi^+} = y_{\xi^+}$ holds in the set $\{\xi \in [0,1) : x_\xi = y_\xi\}$.

Finally, these definitions or conventions are only applicable in the context of representing the sample path of a random process, in which functions are written in "tuple" notation, etc. So, for an ordinary function $f(z)$, the derivative $f'(z)$ retains is usual meaning, as does $\int f(z)dz$.

4.4.4 Discrete Events and Random Times

A *time-dependent* event is a sequence (L_t), where each L_t is an event that depends upon the state of a process at time t. The individual occurrences of such events are not asymptotically "visible," and it makes little sense to consider events of the form $L_{\lfloor \xi T \rfloor}$. We shall represent time-dependent events as follows

- the *indicator process* satisfies $\mathbf{I}_t[L] = 1$ if L_t occurs and $\mathbf{I}_t[L] = 0$ otherwise;
- the *measure process* has increments $\Delta\boldsymbol{\sigma}_{t+1}[L] = \mathbf{I}_t[L]$, and is initially 0, so

$$\boldsymbol{\sigma}_t[L] = \sum_{s=0}^{t-1} \mathbf{I}_s[L],$$

- a measure process is implicitly scaled in time in space, so we have

$$\boldsymbol{\sigma}_\xi[L] = \boldsymbol{\sigma}_{\lfloor \xi T \rfloor}[L]/T,$$

Measure processes will generally be used in conjunction with the Riemann-Stieltjes integral; this will be described below in the next sub-section §4.4.5.

Random Times

A *random time* is simply a random element $\boldsymbol{\tau} \in \{0, \dots, T\}$. In most cases, we will want random times to satisfy some other properties, (i.e. hitting times, adapted times), but, for the purposes of topological representation, all that matters is $\boldsymbol{\tau} \in \{0, \dots, T\}$.

A random time is represented by dividing by the duration of the process. Since this is just one object, no new notation is necessary, we just write $\boldsymbol{\tau}/T \to \tau$ to indicate convergence to a specific $\tau \in [0, 1]$. It is evident that asymptotic equicontinuity ensures that we have

$$\left(\boldsymbol{\tau}/T \to \tau, \mathbf{x}_\tau \to x\right) \implies \mathbf{x}_{\boldsymbol{\tau}} \to x$$

and therefore the state at a random time $\boldsymbol{\tau}$ is completely determined by solutions to $\boldsymbol{\tau}$ and $(\mathbf{x}_\xi)$.

Stopping Times and the Specified Duration

A special case of a random time is a stopping time, meaning that the process terminates at time $\boldsymbol{\tau}$. In this case, we distinguish the *specified duration*, which is the value T, from the actual number of steps before the process stops. By the usual convention, we extend the process for the specified duration in its terminal state, meaning that $\mathbf{x}_t = \mathbf{x}_\tau$ for $\boldsymbol{\tau} \leq t \leq T$. Hence, premature termination does not affect the time-scaling, which depends on the specified duration T, rather than the number of steps which actually take place.

Technically, of course, the specified duration may be random as well. This might occur if the random process is part of some larger random system, and the specified duration would then depend on random parameters which are not part of the actual random process. However, we will not explicitly encounter this situation, and in any case, it is not hard to imagine how to handle this.

4.4.5 Step Decompositions

We will often use measure processes to decompose the expression of a given $(\mathbf{x}_t)$ a sum of increments by only considering steps during which (or immediately after) a particular time-dependent occurs. We shall therefore define a related process

$$\mathbf{x}_t[L] = \sum_{s=0}^{t-1} \mathbf{I}_t[L] \Delta \mathbf{x}_{t+1}, \tag{4.17}$$

so $\mathbf{x}_t[L] = 0$ and the increments are $\Delta \mathbf{x}_t[L] = (\Delta \mathbf{x}_t) \cdot (\Delta \boldsymbol{\sigma}_t[L])$.

Now, the solution to $\mathbf{x}_t[L]$ cannot be determined from solutions to $\mathbf{x}_t$ and $\boldsymbol{\sigma}_t[L]$ using the calculus, because what is being integrated is not $\mathbf{x}_t$ but the increments $\Delta \mathbf{x}_t$, which are typically not well-behaved. As a simple example, if $\Delta \boldsymbol{\sigma}_t[L] = 0$ when t is odd, and $\Delta \mathbf{x}_t = 0$ whenever t is even, then $\mathbf{x}_t[L] = 0$ for all t. On the other hand, if one of these conditions is reversed, then $\mathbf{x}_t[L] = \mathbf{x}_t$. In either case, though, the scaled solutions to both $\mathbf{x}_t$ and $\boldsymbol{\sigma}_t[L]$ will be the same, since whether the increments occur at odd or even times does not affect the asymptotic behavior of a process (provided of course that $\mathbf{x}_t$ is equicontinuous).

The actual value of the process $\mathbf{x}_t[L]$ defined above is not particularly meaningful,

and serves only to provide insight into the increments of $\mathbf{x}_t$ when summed over particular a subset of time steps. In particular, note that

$$\Delta \mathbf{x}_t = \Delta \mathbf{x}_t[L] + \Delta \mathbf{x}_t[\neg L]$$

by definition. We extend this idea as follows:

- a *step decomposition* is a finite or countable set of measure processes $\{\boldsymbol{\sigma}_t[L_j]\}_{j\in J}$ such that $\sum_{j\in J} \Delta \boldsymbol{\sigma}_t[L_j] = 1$ for all t.

For a step decomposition, then, the events L_j are mutually exclusive, and exactly one of them must occur each step.

Proposition 4.4.6. *If $(\mathbf{x}_t)$ is Lipschitz, and $\{\boldsymbol{\sigma}_t[L_j]\}_{j\in J}$ is a finite step decomposition, then the differential constraint*

$$dx_\xi = \sum_{j\in J} dx_\xi[L_j]$$

holds u.a.

Proof. Since $(\mathbf{x}_t)$ is Lipschitz, then clearly so is each $(\mathbf{x}_t[L_j])$. Hence, the solution to each will also be Lipschitz, and we also have the constraint $(x_\xi) = \sum_{j\in J}(x_\xi[L_j])$ which leads to the above differential constraint. □

We note that the assumptions in this proposition can be weakened considerably; for example, the step decomposition can be allowed to be countable, provided that it is also summable. The Lipschitz assumption can also be weakened, but there are some technicalities involved, since if $(\mathbf{x}_t)$ is not Lipschitz, then it is possible that $(\mathbf{x}_t[L])$ is not asymptotically equicontinuous.

Partial Derivative Notation

As noted, we are generally interested only in processes of the form $(\mathbf{x}_t[L])$ for differential reasons, and the actual value of this process is not meaningful. Accordingly, we will

abbreviate by using partial derivative notation, as in

$$\frac{\partial x_\xi}{\partial \sigma_\xi[L]} = \frac{dx_\xi[L]}{d\xi}.$$

We emphasize that this is just notation; there are no actual partial derivatives. The intuition is that is $\frac{\partial x_\xi}{\partial \sigma_\xi[L]}$ is the "part" of dx_ξ which coincides with increments of the measure process $\boldsymbol{\sigma}_t[L]$. It is perhaps possible to construe this as a partial derivative in some rigorous sense, but this is not necessary for our purposes, and we will leave this as simply a notational abbreviation.

4.5 Martingale Concentration

The tools we have developed thus far have relied entirely on non-random properties of various topological spaces. These tools yield what we have called *structural constraints*, which generally depend on the specification of a random element or process, and not on the actual probabilistic behavior. As a result, structural constraints will usually hold u.a. (ultimately always) rather than just w.h.p.[17]

In order to usefully apply these structural results to random structures, we must combine them with *probabilistic constraints* that hold strictly w.h.p. and not u.a. Indeed, without any additional constraints, the results discussed in the previous sections essentially constitute nothing more than some elementary and well-known facts about topologies and function spaces.

In this section, we add the "missing ingredient" to our recipe, in the form of Azuma's classical martingale concentration inequality. For most of our applications, Azuma's inequality is in fact the only "w.h.p. constraint" we shall need. As we shall see, this single probabilistic constraint, when combined with the non-random tools developed in the previous sections, results in a fairly powerful and general method for solving discrete random processes.

[17] Of course, it is not necessary for a structural constraint to hold u.a., and if the if the assumptions for such a constraint are only satisfied w.h.p. rather than u.a., then the consequences will also only hold w.h.p.

4.5.1 Azuma's Inequality and Doob Decomposition

We begin by restating Azuma's martingale concentration inequality (corollary 2.3.4 from chapter 2) in an asymptotic form consistent with the definitions and notations we have presented thus far. This will yield a constraint that holds w.e.h.p., and we recall that the w.e.h.p. condition specifies an exponential rate of convergence in probability. We will use the specified duration T for this purpose, and hence we require the assumption that the specified duration tends to ∞.

Proposition 4.5.1 (Azuma's Inequality)**.** *Let* $(\mathbf{x}_t)$ *be a martingale in* $\mathbb{R}$ *for which:*

- *the initial state is* $\mathbf{x}_0 = 0$ *and the specified duration satisfies* $T \to \infty$*;*
- *the Lipschitz condition* $\sup\{|\Delta \mathbf{x}_t| : 1 \leq t \leq T\} < c/T$ *holds u.a. for some fixed constant* c*.*

Then $\|(\mathbf{x}_\xi)\|_\infty \to 0$ *holds* T*-w.e.h.p.*

Proof. Note that the Lipschitz condition, along with the assumption that the specified duration $T \to \infty$, ensures that $(\mathbf{x}_t)$ is asymptotically equicontinuous. It thus suffices to prove that $\mathbf{x}_\xi = \mathbf{x}_{\lfloor \xi T \rfloor} \to 0$ holds T-w.e.h.p. for any fixed $\xi > 0$.

Azuma's inequality (corollary 2.3.4 in chapter 2) assures us that, for any such $\xi > 0$, and any $\epsilon > 0$, there exists a constant C such that

$$\mathsf{P}\left[\left|\mathbf{x}_{\lfloor \xi T \rfloor}\right| > \epsilon \lfloor \xi T \rfloor\right] \leq e^{-C\lfloor \xi T \rfloor} = e^{-\Omega(T)},$$

which is exponentially small with respect to $T \to \infty$. Since this is true for any $\epsilon > 0$, then $\mathbf{x}_{\lfloor \xi T \rfloor} \to 0$ holds T-w.e.h.p., and the proof is complete. □

Some comments:

1. The Lipschitz condition must hold u.a. and not just w.h.p. While there are ways to weaken this assumption, we will not need to do so. Briefly, the problem with a

Lipschitz condition that only holds w.h.p. is not that the behavior of the process changes dramatically, but rather that the small possibility of an enormous increment size throws off the computation of expected increments.

2. More generally, the strict Lipschitz condition can be replaced with a number of weaker conditions, and exponentially high concentration can still be established, but the Lipschitz condition will hold in our applications.

3. It is also possible to achieve weaker guarantees in probability (e.g. w.p.h.p.), if the bound on increments if of the form $|\Delta \mathbf{x}_t| < f(T)$ for some non-constant function $f(T) \to \infty$ as $T \to \infty$. In this situation, the Lipschitz condition will not immediately imply equicontinuity, so some additional work is required, and we omit this argument since it is not necessary for our purposes.

Intuitively, Azuma's inequality gives us one additional constraint on the solution space of any random process by assuring us that any Lipschitz martingale is asymptotically equivalent to a constant function. Of course, in general our random processes will not all be martingales. However, by using a simple technique called *Doob decomposition* (simple, at least, in this discrete setting), we may extract a martingale from any discrete random process (with finite expectation) as follows.

Definition 4.5.1. For an $\mathbb{R}$-valued process $(\mathbf{x}_t)$ which satisfies $\mathsf{E}[|\mathbf{x}_t|] < \infty$ for all t:

- the *predictable component* is the process

$$\mathfrak{E}\mathbf{x}_t = \mathbf{x}_0 + \sum_{s=0}^{t-1} \mathsf{E}_s[\Delta \mathbf{x}_{s+1}], \tag{4.18}$$

- the *martingale component* is the process

$$\mathfrak{M}\mathbf{x}_t = \sum_{s=0}^{t-1} \Delta \mathbf{x}_{s+1} - \mathsf{E}_s[\Delta \mathbf{x}_{s+1}], \tag{4.19}$$

- the *Doob decomposition* is $\mathbf{x}_t = \mathfrak{M}\mathbf{x}_t + \mathfrak{E}\mathbf{x}_t$

Note that the martingale component $(\mathfrak{M}\mathbf{x}_t)$ is clearly, as a name suggests, a martingale, and the predictable component satisfies

$$\Delta\mathfrak{E}\mathbf{x}_{t+1} = \mathsf{E}_t[\Delta\mathbf{x}_{t+1}] = \mathsf{E}_t[\Delta\mathfrak{E}\mathbf{x}_{t+1}],$$

and is predictable in the sense that the value of $\mathfrak{E}\mathbf{x}_{t+1}$ can be determined at time t.[18]

We easily obtain the following corollary to proposition 4.5.1.

Corollary 4.5.2. *Let $(\mathbf{x}_t)$ be a Lipschitz process with duration $T \to \infty$. For any fixed function (x_ξ):*

$$(\mathfrak{E}\mathbf{x}_\xi) \to (x_\xi) \implies (\mathbf{x}_\xi) \to (x_\xi) \quad T\text{-w.e.h.p.}$$

Proof. Since expected increments are clearly bounded by the maximum size of true increments, then if $(\mathbf{x}_t)$ is Lipschitz, then so is $(\mathfrak{E}\mathbf{x}_t)$, and therefore $(\mathfrak{M}_t) = (\mathbf{x}_t) - (\mathfrak{E}\mathbf{x}_t)$ is a Lipschitz martingale. The above implication follows from Azuma's equality. □

The power of this result comes from the fact that any Lipschitz process can thus be solved w.e.h.p. by determining the expected increments $\Delta\mathfrak{E}\mathbf{x}_{t+1} = \mathsf{E}_t[\Delta\mathbf{x}_{t+1}]$ rather than the actual increments $\Delta\mathbf{x}_{t+1}$. Unlike the actual increments, the expected increments are often "well-behaved," and thus easier to predict in general.

4.5.2 Differential Equations in One Dimension

We now give a simple example of how martingale concentration can be combined with the methods in the previous section to solve a one-dimensional process using differential equations.

Theorem 4.5.3. *Let $(\mathbf{x}_t)$ be a Lipschitz process with duration $T \to \infty$, and assume that, for a continuous function f, expected increments satisfy $\mathsf{E}_t[\Delta\mathbf{x}_{t+1}] = f(\mathbf{x}_t)/T$ for all t. Then every T-w.e.h.p. solution to $(\mathbf{x}_\xi)$ must satisfy*

$$dx_\xi = f(x_\xi)d\xi$$

[18] At a technical level, we are assuming that $(\mathbf{x}_t)$ is adapted to some filtration $(\mathcal{F}_t)$, and thus the martingale component is an $(\mathcal{F}_t)$-adapted martingale, while the predictable component is $(\mathcal{F}_t)$-predictable.

for all $\xi \in [0,1)$.

Proof. Since f is continuous then for any (x_ξ), assuming w.l.o.g. that $\mathbf{x}_0 = 0$, we have

$$\begin{aligned}
(\mathbf{x}_\xi) \to (x_\xi) &\implies (f(\mathbf{x}_\xi)) \to (f(x_\xi)) && \text{u.a., by continuity} \\
&\implies (T \cdot \Delta\mathfrak{E}\mathbf{x}_{\lfloor \xi T \rfloor}) \to (f(x_\xi)) && \text{u.a., by assumption} \\
&\implies (\mathfrak{E}\mathbf{x}_\xi) \to (\textstyle\int f(x_\xi)) && \text{u.a., by corollary 4.4.4} \\
&\implies (\mathbf{x}_\xi) \to (\textstyle\int f(x_\xi)) && \text{w.e.h.p., by martingale concentration.}
\end{aligned}$$

Since the solution space $\mathcal{C}([0,1],\mathbb{R})$ is Hausdorff, it follows that any (x_ξ) which does not satisfy this integral equation is a non-solution, and since f is continuous, this is equivalent to the differential equation $dx_\xi = f(x_\xi)d\xi$ for all $\xi \in [0,1)$. □

Note, though, that this corollary does not necessarily imply that there exists a unique (i.e. strong) solution. Indeed, if this differential equation has multiple solutions, then all we can conclude is that any solution to process $(\mathbf{x}_\xi)$ must satisfy $dx_\xi = f(x_\xi)d\xi$. In order to find a unique solution, some additional assumptions are required.

Corollary 4.5.4. *Assume the conditions of theorem 4.5.3 are satisfied, and in addition:*

1. *there exists a fixed* $x \in \mathbb{R}$ *such that* $\mathbf{x}_0 \to x$*;*
2. (x_ξ) *is the unique solution to* $dx_\xi = f(x_\xi)d\xi$ *with initial condition* $x_0 = x$*.*

Then $(\mathbf{x}_\xi) \to (x_\xi)$ *w.e.h.p.*

Proof. (x_ξ) is clearly the unique weak solution in this case, and since $(\mathbf{x}_\xi)$ has compact solution space, this weak solution must be a strong solution. □

It is worth noting that all of the above convergences also hold uniformly, since the processes involved are Lipschitz and therefore asymptotically equicontinuous. At a technical level, there is perhaps an additional step somewhere in the proof where pointwise convergence is turned into uniform convergence. However, since the Lipschitz condition is also necessary for martingale concentration, then this is not a concern.

4.5.3 The General Differential Equations Method

We now discuss the method of differential equations in full generality. Unlike the simple one-dimensional example given above, the general method does not consist of one single theorem which immediately translates certain conditions about a random process into a strong solution. Instead, what we have is a general method for translating attributes of a random process to differential solution constraints. Hence, a more appropriate general theorem statement might be the following.

Theorem 4.5.5. *Let* $(\mathbf{x}_t)$ *and* $(\mathbf{y}_t)$ *be* $\mathbb{R}$*-valued process, assume* $(\mathbf{x}_t)$ *is Lipschitz and* $(\mathbf{y}_t)$ *is asymptotically equicontinuous, and assume* $T \to \infty$*.*

Then, if $\mathsf{E}_t[\Delta \mathbf{x}_{t+1}] = \mathbf{y}_t/T$ *for all* t*, the constraint*

$$dx_\xi = y_\xi d\xi$$

holds T*-w.e.h.p. for every* $\xi \in [0, 1)$

Proof. It is immediate by corollary 4.4.4 that the predictable component of $(\mathbf{x}_t)$ is completely determined by $(\mathbf{y}_t)$, and due to the Lipschitz condition on $(\mathbf{x}_t)$, the martingale component converges w.e.h.p. to 0. It follows that every w.e.h.p. solution must satisfy

$$x_\xi = x_0 + \int_{\zeta=0}^{\xi} y_\zeta d\zeta$$

for all ξ, which yields the above w.e.h.p. differential constraint. □

In applications, differential constraints of this sort can be used in a number of ways, and it is not necessary to compute a complete system of differential equations all in one shot. The basic techniques we have developed are sufficiently malleable to handle a reasonably broad set of circumstances, and we now briefly discuss some generalizations of the one-dimensional example in theorem 4.5.3.

Multiple Dimensions

The argument in theorem 4.5.3 generalizes quite easily to multiple or even (countably) infinite dimensions. In fact, for finite dimensions there is practically nothing else to prove. If the state space is $\mathbb{R}^\omega$, there are issues related to the topology of the state space, and what qualifies as a continuous function $\mathbb{R}^\omega \to \mathbb{R}$ or $\mathbb{R}^\omega \to \mathbb{R}^\omega$, but these are not difficult to resolve.

For our applications, the state space will be the set $\Phi \subseteq \mathbb{R}^\omega$ of distributions on $\mathbb{Z}^*$, corresponding to the degree or residual distribution of a graph. The particulars of this situation will be discussed in detail in the next chapter, but, modulo certain technicalities, nothing is much different.

Of course, the resulting system of differential equations will be infinite dimensional, and finding a solution is not always easy. But, this is where the difficulty belongs, rather than in the task of proving that these differential equations govern the solution of our random process.

Transitions and Terminations

In the above example, the same differential constraint was applicable for the duration of the process, but obviously this is without loss of generality, and a differential constraint can hold on some sub-interval $[\zeta_1, \zeta_2]$. A more common situation is if a differential constraint only holds before a random time $\boldsymbol{\tau}$, at which either the process terminates, or a transition to a different, qualitatively different phase of the process takes place. In most cases, transitions of this sort can be dealt with rather smoothly using solution constraints, and without needing to tangle with low-level details of the sample path.

Consider a very simple example: we have a (Lipschitz) process with state space $[0, 1]$, which terminates at the first time $\boldsymbol{\tau}$ for which $\mathbf{x}_\tau = 0$, and such that $\mathsf{E}_t[\Delta\mathbf{x}_{t+1}] = f(\mathbf{x}_t)$ holds until then. Now, the termination condition cannot be detected asymptotically, since even if $\mathbf{x}_\xi \to x_\xi = 0$, it may be the case that the actual state never reaches 0, but just remains small. On the other hand, if $\mathbf{x}_\xi \to x_\xi > 0$, then we must actually have $\mathbf{x}_\zeta > 0$ for

$\zeta \in (\xi - \epsilon, \xi + \epsilon)$ by equicontinuity.

This would yield to the following system of constraints:

1. if $\xi < \tau$ then $dx_\xi = f(x_\xi)d\xi$ w.e.h.p.;
2. if $x_\xi > 0$ than $\tau \neq \xi$ u.a.

Note that the first of these holds only w.e.h.p., since it involves martingale concentration, while the second holds u.a., since this is a structural constraint which only depends on equicontinuity.

If this differential equation is uniquely solvable, and if the solution reaches $x_\xi = 0$ at some $\xi < 1$, then, as noted, we cannot yet conclude that $\tau = \xi$. But, if $f(0) < 0$, then we can indeed draw this conclusion, since if $\tau > \xi$, we would have $dx_\xi < 0$, which would imply that the state must become negative, which is impossible. In this case, the above constraints would have a unique solution.

More generally, the topological representation allows us deal with transitional behavior of a discrete process in terms of the possible ways that a continuous function with bounded variation can cross from one region of the state space to another, rather than the ways a discrete sample path can do the same. This allows us to use various standard tools related to continuity and almost everywhere differentiability. For example, if $x_{\zeta_1} = 0$ and $x_{\zeta_2} > 0$ for $\zeta_2 > \zeta_1$ then the following must hold:

- there must be some $\xi \in (\zeta_1, \zeta_2)$ such that $dx_\xi > 0$ (i.e. the mean value theorem);
- there must be some $\xi_0 \in [\zeta_1, \zeta_2)$ such that $x_{\xi_0} = 0$ and $x_\xi > 0$ for $\xi \in (\xi_0, \xi_0 + \epsilon)$ (since the set $\{\xi : x_\xi > 0\}$ is open).

Due to these and other similarly basic facts, analyzing transitions using continuous functions is often much simpler than dealing with the actual sample path.

Non-Homogeneous Behavior

In order for the differential equations technique to work correctly, some sort of Markov property must hold, which allows expected increments to be determined from the current state and not the distant history. However, it need not be a strict Markov condition.

We may easily imagine a situation where, in addition to $\mathbf{x}_t \in \mathbb{R}$, there are two possible discrete states, and depending on the which of these states the process is in at a given time, the expected increment behaves differently. Using measure processes, we may express this situation as

$$dx_\xi = f_1(x_\xi)d\sigma_\xi[1] + f_2(x_\xi)d\sigma_\xi[2].$$

This of course introduces two additional functions to the solution space, namely $(\sigma_\xi[1])$ and $(\sigma_\xi[2])$, and since there are only two discrete states, we immediately have the constraint $d\sigma_\xi[1]+d\sigma_\xi[2] = d\xi$. This leaves one remaining degree of freedom, which must be eliminated by other means. For example, if the discrete states behave like a 2-state Markov chain, we can use the transition probabilities to determine the relative frequency of each state, etc.

4.5.4 Unique Solutions to Differential Equations

The fact that a certain set of differential constraints hold w.e.h.p. does not imply that these constraints or uniquely solvable, nor should it. Indeed, the question of whether a given solution is unique is of a different nature than the question of whether a these differential equations characterize the asymptotic behavior of a random process. In fact, even if we cannot find a unique solution, we may yet gain some information about the asymptotic behavior of $(\mathbf{x}_\xi)$ from the set of all such solutions, which in this case is the minimal solution space. So, for example, if every solution satisfies $x_\zeta = y$ at a given ζ, then it follows that $\mathbf{x}_\zeta \to y$ w.e.h.p.

The processes we study will, in most cases, have unique solutions, and we now review the basic conditions which imply that an initial value problem is uniquely solvable. For finite dimensional processes, the basic requirement is a Lipschitz condition.

Our notation for dealing with $\mathbb{R}^k$ is somewhat "backwards" in that we shall index coordinates in "function" notation, so $x = (x(1), \ldots, x(k))$, and an element in the function space $\mathcal{M}([0,1], \mathbb{R}^k)$ is denoted by $(x_\xi) = (x_\xi(1), \ldots, x_\xi(k))$.

An actual "function" $F : \mathbb{R}^k \times [0,1] \to \mathbb{R}^k$, though, is written as usual, so

$$F(x, \xi) = (F_i(x, \xi), \ldots, F_k(x, \xi)),$$

and integrals are performed coordinatewise, in the natural way. Also, in the space $\mathbb{R}^k$, we let $\|(\| x)$ denote the standard Euclidean norm.

Theorem 4.5.6 (Picard-Lindelöf)**.** *Let F be a continuous function $\mathbb{R}^k \times [0,1] \to \mathbb{R}^k$, let $x \in \mathbb{R}^k$ and assume that, for fixed constants $\epsilon, C > 0$, and an open neighborhood $B \ni x$, the Lipschitz condition*

$$\|F(y, \xi) - F(z, \xi)\| < C \|y - z\|$$

holds for all $y, z \in B$ and all $\xi \in [0, \epsilon]$.

Then, there exists $\delta > 0$ such that the integral equation

$$x_\xi = x + \int_{\zeta=0}^{\xi} F(x_\zeta, \zeta) d\zeta \tag{4.20}$$

has exactly one solution for $\xi \in [0, \delta]$.

Proof. For a complete proof, see, e.g., [21]. Briefly, we may assume w.l.o.g. that:

- the initial state is $x = (0, \ldots, 0)$, the neighborhood B is all of $\mathbb{R}^k$, and
- both constants are $C = \epsilon = 1$, and also $\|F(y, \xi)\| < 1$ for all y, ξ.

In this case, if we let $X = \{x \in \mathbb{R}^k : \|x\| \leq 1\}$, then the set of possible solutions to (4.20) is contained in

$$\mathcal{X} = \{(x_\xi) \in \mathcal{C}([0,1], X) : \|x_\xi - x_\zeta\| \leq |\xi - \zeta|\},$$

and by Arzelà-Ascoli , this set is compact in $\mathcal{C}([0,1], X)$ with respect to the uniform (i.e. ℓ_∞) norm.

For any $(x_\xi), (y_\xi) \in \mathcal{X}$, we have

$$\left\| \int F(x_\xi, \xi) - \int F(y_\xi, \xi) \right\|_\infty \leq \int_{\xi=0}^{1} \|F(x_\xi, \xi) - F(y_\xi, \xi)\| < \|(x_\xi) - (y_\xi)\|_\infty .$$

Hence, the mapping $(x_\xi) \mapsto (\int F(x_\xi, \xi))$ is a contraction mapping and thus there exists exactly one fixed point. □

There are obvious trivial generalizations. For instance, it suffices for F to be piecewise continuous in the time coordinate, since the only real need for continuity is to ensure that the integral in (4.20) is well-defined. The generalization to infinite dimensions is non-trivial, and there is no simple way to characterize unique solvability for infinite systems of differential equations. However, as we discuss in the next chapter, the infinite systems of differential equations we encounter are easily shown to have unique solutions by considering finite sub-systems.

Chapter 5

Differential Equations and Random Graph Algorithms

In the previous chapter, we developed a suite of tools geared towards the asymptotic analysis of discrete random structures and processes. In abstract terms, the fundamental idea is *topological representation*, which involves mapping a random discrete into a topological space in such a way that the pertinent asymptotic information can be expressed via convergences and separations. The problem then becomes one of finding a w.h.p. point, and we describe this problem using the "constraint satisfaction" metaphor: a w.h.p. limit is a *solution*, and potential solutions are ruled out by *solution constraints*. This topological abstraction provides the technical foundation for the differential equations method, in which case the topological representation is the scaled sample path of a discrete random process, and the solution constraints take the form of differential equations.

In this chapter, we describe the specific topological representation of various structures which arise from algorithmic study of the configuration model. The essential random structure, which was introduced in chapter 3 is the *endpoint removal process*, which is a random descending chain of subsets $(\mathbf{A}_0 \supseteq \cdots \supseteq \mathbf{A}_T)$ of the endpoint set $\mathbf{A}_0 = A$ for a given endpoint partition $H = (A, V)$.

The topological representation of a single endpoint partition H was already described in chapter 3, in terms of either degree distribution λ_H, or the residual distribution μ_H. Accordingly, an endpoint removal process will be represented by the corresponding processes $(\boldsymbol{\lambda}_t)$ and $(\boldsymbol{\mu}_t)$, for which the state space is the space $\Phi = \mathrm{Dist}(\mathbb{Z}^*)$ of distributions on $\mathbb{Z}^*$, and our present objective is to go over certain technicalities and develop some tools to analyze processes of this kind.

Chapter Organization

We begin in section 5.1 by describing the topological representation of an endpoint removal process. The "increment" of a removal process is the endpoint $\mathbf{a}_t$ which is removed from $\mathbf{A}_{t-1}$ to yield $\mathbf{a}_t$, and in section 5.2, we describe certain step decompositions which we shall use to relate properties of this endpoint to the increments of the degree and residual distributions.

In section 5.3 we analyze an endpoint removal process in which every endpoint is removed uniformly at random. While such a process is not particularly interesting in terms of random graph properties, the solution can be used in various ways help analyze other algorithms which combine random and deterministic endpoint removal.

In section 5.4, we discuss the use of the probability generating function (p.g.f.) as a tool for manipulating distributions on the set $\mathbb{Z}^*$.

5.1 The Endpoint Removal Process

The fundamental random structure in our algorithmic analysis of the configuration model is the *endpoint removal process*, which was introduced in section 3.2 of chapter 3. Given an endpoint partition (A, V):

- a *removal process* is a random descending chain of subsets $(A = \mathbf{A}_0 \supseteq \cdots \supseteq \mathbf{A}_T)$;
- such a process is *simple* if exactly one endpoint $\mathbf{a}_{t+1}$ is removed from $\mathbf{A}_t$ each step:

$$\mathbf{A}_{t+1} = \mathbf{A}_t \backslash \{\mathbf{a}_{t+1}\}.$$

There are many variations and alternate interpretations of a removal process, which we will not get into in this chapter; a brief summary is given below. For now, we will assume that we are "given" an endpoint removal process, and discuss techniques for analyzing this process asymptotically.

Recall that, as discussed in chapter 3, our asymptotic parametrization of the configuration model involves convergence of the residual distribution $\mu_H \to H$. This convergence,

along with the assumption along with the assumption that $m(H) = |A| \to \infty$ is abbreviated by $H = \mathcal{H}(\mu)$. Our standard regularity assumptions for an endpoint removal process are as follows.

Definition 5.1.1. A removal process is *regular* if:

1. the initial state satisfies $H = \mathcal{H}(\mu)$ for some $\mu \in \Phi$;
2. the initial degree distribution converges $\lambda_H \to \lambda$ to some $\lambda \in \Phi$ with $\lambda(0) < 1$;
3. the process is simple.

The reason for the first assumption is obvious, and the other two are more or less without loss of generality. Since $\lambda_H(0)$ is the fraction of isolated vertices, which have no meaningful effect on the structure of a graph, then for practical purposes we might even assume that $\lambda(0) = 0$ if we choose and then rescale whatever results as need.

The algorithms we analyze will generally not satisfy the simplicity assumption in their literal form; for example, removing an edge from a configuration involves removing two endpoints at once. However, such a removal process can generally be made simple by decomposing *iterations* of an algorithm into individual endpoint removal *steps*, and dealing with simple processes is preferable, since the specified duration is then equal to the number of endpoints $t = m = |A|$.

Endpoint Removal and the Algorithmic Method

We now briefly summarize the key developments which lead to the definition of an endpoint removal process. First, while we are using the term "graph" for general discussion, the objects we are dealing with are in fact configurations, each of which is a triple $G = (A, V, E)$ of endpoints, vertices, and edges. An endpoint partition $\mathbb{H}(G) = (A, V)$ is just a configuration with no edges, and given any $H = (A, V) \in \mathbb{H}$, the random configuration $\boldsymbol{\mathcal{G}}(H) = (A, V, \mathbf{E})$ is constructed by choosing the edge set $\mathbf{E}$ uniformly at random from the set $\mathbb{E}(A)$ of perfect matchings of A.

Our definition of a *graph algorithm* consists simply of a transition function $\pi : \mathbb{G} \to \mathbb{G}$ (along with a set of terminal states $\mathbb{G}_T$), and the *execution path* of such an algorithm is the $\mathbb{G}$-valued random process, where the initial state is a random graph $\boldsymbol{\mathcal{G}}(H)$, and such that $\mathbf{G}_{t+1} = \pi(\mathbf{G}_t)$. Our techniques require that the transition function π is *uniformity preserving*, meaning that $\pi(\boldsymbol{\mathcal{G}}(H))$ will be uniformly distributed conditional on its endpoint partition $\mathbb{H}(\pi(\boldsymbol{\mathcal{G}}(H)))$.

In this case, the distribution of $\mathbf{G}_t$ at any time during the execution path will also by $\mathbb{H}$-conditionally uniform. It follows that the *observable process* $(\mathbf{H}_t)$, which is defined by $\mathbf{H}_t = \mathbb{H}(\mathbf{G}_t)$, will be an $\mathbb{H}$-valued Markov chain. The algorithms we analyze will typically remove edges from a configuration without affecting the vertex set, and hence the observable process will be of the form $\mathbf{H}_t = (\mathbf{A}_t, V)$, where $(\mathbf{A}_t)$ is an endpoint removal process.

5.1.1 Basic Notation and Topological Representation

The analysis of any endpoint removal process is based on random sequence of endpoint partitions $(\mathbf{H}_t) = (\mathbf{A}_t, V)$, but since the vertex set remains constant, the only substantive part of the state is the endpoint set $\mathbf{A}_t$. Accordingly, we will generally abbreviate our notation by using $\mathbf{A}_t$ in place of $\mathbf{H}_t$, and we will further abbreviate by denoting "anything" that depends on the state at time t by "$\mathbf{anything}_t$", most notably:

- $\mathbf{deg}_t(v) = \deg_{\mathbf{A}_t}(v)$ is the degree of a vertex $v \in V$;
- $\mathbf{res}_t(a) = \mathrm{res}_{\mathbf{A}_t}(a)$ is the residual degree of an endpoint $a \in \mathbf{A}_t$;
- $\mathbf{m}_t = |\mathbf{A}_t|$ is the number of remaining endpoints (the number of vertices is always $n = |V|$);
- $\boldsymbol{\lambda}_t = \lambda_{\mathbf{A}_t}$ is the degree distribution;
- $\boldsymbol{\mu}_t = \mu_{\mathbf{A}_t}$ is the residual distribution;

Degrees and residual degrees were discussed at length in chapter 3, so we will not go into details about what these mean.

It is worth noting, though, that the residual distribution $\boldsymbol{\mu}_t$ only counts endpoints which belong to $\mathbf{A}_t$, so we have

$$\boldsymbol{\mu}_t(i) = \frac{|\{a \in \mathbf{A}_t : \mathbf{res}_t(a) = i\}|}{\mathbf{m}_t},$$

assuming, of course, that $\mathbf{m}_t > 0$. In a simple removal process, only one endpoint is removed each step, so $\mathbf{m}_t = m - t$, where $m = |A|$ is the size of the initial endpoint set.

We say an endpoint is $a \in A$ is *unexposed* at time t if it belongs to $\mathbf{A}_t$, and otherwise a is *exposed*; "exposed" is therefore synonymous with "removed." Accordingly, any other structure related to either $\mathbf{A}_t$ or $\mathbf{H}_t$ is similarly called unexposed (e.g. the unexposed degree distribution is $\boldsymbol{\lambda}_t$).

Topological Representation

Representing a discrete random process topologically requires scaling both the time and space coordinates. In this case, the space coordinate is already scaled via the degree or residual distribution. For a simple process, the specified duration is just $T = m$, and so we scale the time coordinate by

$$\boldsymbol{\lambda}_\xi = \boldsymbol{\lambda}_{\lfloor \xi T \rfloor} = \boldsymbol{\lambda}_{\lfloor \xi m \rfloor},$$

and similarly for the residual distribution.

$(\boldsymbol{\lambda}_\xi)$ is therefore a random element in the space $\mathcal{M}([0,1], \Phi)$ of mappings from $[0,1]$ to $\Phi = \mathrm{Dist}(\mathbb{Z}^*)$. By default, this space is endowed with product topology (over both coordinates), so a convergence $(\boldsymbol{\lambda}_\xi) \to (\lambda_\xi)$ indicates that $\boldsymbol{\lambda}_\xi(i) \to \lambda_\xi(i)$ for every fixed pair of coordinates $\xi, \in [0,1]$ and $i \in \mathbb{Z}^*$.

As usual, all of these structures are understood to be asymptotic, so $\boldsymbol{\lambda}_\xi = \boldsymbol{\lambda}_{\eta,\xi}$, and this convergence occurs in the limit as $\eta \to \infty$. Moreover, in this context, all w.e.h.p. and w.p.h.p. claims are with respect to the number of endpoints m. Since this is the same as the specified duration T, then the version of Azuma's inequality in proposition 4.5.1 from the previous chapter yields convergence which holds m-w.e.h.p.

When working with the residual distribution, we will restrict our attention to the interval $\xi \in [0, 1)$. The trivial reason for this is that the residual distribution is not well-defined in the terminal state, since $|\mathbf{A}_m| = 0$. We could obviously get around this in a number of ways; a more significant issue is that, due to the dependence on $1/\mathbf{m}_t$, the residual distribution behaves poorly for $t = m - o(m)$. For both of these reasons, we only define the function $(\boldsymbol{\mu}_\xi)$ for $\xi \in [0, 1)$.

We finally note that, although both $\boldsymbol{\lambda}_\xi$ and $\boldsymbol{\mu}_\xi$ (for fixed $\xi < 1$) belong to the set Φ of distributions on $\mathbb{Z}^*$, this set is not closed. Hence, at the outset, we cannot rule out the possibility that $\boldsymbol{\lambda}_\xi \to \lambda_\xi$ for some $\lambda_\xi \notin \Phi$. The representation space for $\boldsymbol{\lambda}_\xi$ should thus be considered simply the countable product space $\mathbb{R}^\omega$, and accordingly, the representation space for $(\boldsymbol{\lambda}_\xi)$ is $\mathcal{M}([0, 1], \mathbb{R}^\omega)$.

5.1.2 Consequences of the Regularity Assumptions

We now explore some basic consequences of the regularity assumptions defined above. Perhaps the most significant such consequence is that $(\boldsymbol{\lambda}_\xi)$ is *uniformly summable*, which, as defined in section 4.3 of the previous chapter, means that pointwise convergence implies uniform convergence in ℓ_1:

$$(\boldsymbol{\lambda}_\xi) \to (\lambda_\xi) \implies (\boldsymbol{\lambda}_\xi) \xrightarrow{\infty,\ell_1} (\lambda_\xi).$$

This allows us to use the product topology (more precisely, the product-product topology) for the purposes of establishing convergence, and the stronger uniform-ℓ_1 topology in order to deduce consequences of this convergence.

Theorem 5.1.1. *For any regular endpoint removal process:*

1. *each coordinate process $(\boldsymbol{\lambda}_t(i))$ is Lipschitz;*
2. *$(\boldsymbol{\lambda}_\xi)$ is uniformly summable, and has solution space in some compact subset of $\mathcal{C}([0, 1], \Phi)$;*
3. *if $M_k(\boldsymbol{\lambda}_0) \to M_k(\lambda) < \infty$, then the process $(M_k(\boldsymbol{\lambda}_\xi))$ is asymptotically equicontinuous;*
4. *the above hold for $(\boldsymbol{\mu}_\xi)$ when restricted to an interval $\xi \in [0, 1 - \epsilon]$;*

Proof. First, for the degree distribution, it is clear that $|\Delta\boldsymbol{\lambda}_t(i)| \leq 1/n = O(1/m)$, since at most one vertex changes degree each step. This in turn implies that each coordinate is asymptotically equicontinuous.

For uniform summability, it suffices to show that for any $\epsilon > 0$, there exists J such that $\sup\{\sum_{i>J} |\boldsymbol{\lambda}_\xi(i)| : \xi \in [0,1]\} < \epsilon$ u.a. In this case, since endpoints are only being removed as time passes, then the degree of any vertex is non-increasing. Hence

$$\sup\left\{\sum_{i>J} |\boldsymbol{\lambda}_\xi(i)| : \xi \in [0,1]\right\} = \sum_{i>J} |\boldsymbol{\lambda}_0(i)|\,,$$

and uniform summability follows from the fact that the initial degree distribution is summable by assumption.

For higher moments, the assumption $M_k(\lambda) \to M_k(\lambda_H) < \infty$ similarly implies that, for any $\epsilon > 0$, we can choose J such that $\sum_{i>J} i^k \lambda_H(i) \to \sum_{i>J} (i)_k \lambda(i) < \epsilon$. Since the mapping $i \mapsto (i)_k$ is non-decreasing,[1] asymptotic equicontinuity of higher moments follows by an identical argument to uniform summability of $(\boldsymbol{\lambda}_\xi)$.

Let us now restrict our attention to an interval $\xi \in [0, 1-\epsilon]$, and recall the equation relating the degree and residual distribution from §3.1.2 of chapter 3:

$$\boldsymbol{\mu}_t(i) = \frac{(i+1)\boldsymbol{\lambda}_t(i+1)}{M(\boldsymbol{\lambda}_t)}. \tag{5.1}$$

Since exactly one endpoint is removed each step then we have $\mathbf{m}_t = m - t$, and hence the average degree $M(\boldsymbol{\lambda}_t) = \mathbf{m}_t/n$ is a Lipschitz process. Moreover, on the interval $t/m \in [0, 1-\epsilon]$, $M(\boldsymbol{\lambda}_t)$ is bounded away from 0, and hence its inverse is also a Lipschitz process, and therefore each $(\boldsymbol{\mu}_t(i))$ is a product of two Lipschitz processes which must also be Lipschitz.

Uniform summability, and equicontinuity of higher moments, can now be established using identical arguments to above, again based on the fact that $M(\boldsymbol{\lambda}_t)$ is bounded away from 0. □

[1] Here $(i)_k$ denotes the falling factorial $i(i-1)\cdots(i-k+1)$, and $M_k(\lambda) = \sum_i (i)_k \lambda(i)$ denotes the k'th factorial moment, as discussed in chapter 3.

Due to this theorem, the residual distribution can be understood as just a change of variables from the degree distribution, and we now summarize some key formulas related to this change of variables.

Corollary 5.1.2. *For any regular endpoint removal process:*

1. *the average degree* $(M(\boldsymbol{\lambda}_t)) = (\mathbf{m}_t/n)$ *is Lipschitz, and has unique solution*

$$M(\lambda_\xi) = M(\lambda)(1-\xi), \tag{5.2}$$

2. *for any* $\xi \in [0,1)$, $\boldsymbol{\mu}_\xi$ *and* $\boldsymbol{\lambda}_\xi$ *completely determine each other by*

$$\mu_\xi(i) = \frac{(i+1)\lambda_\xi(i)}{M(\lambda)(1-\xi)}. \tag{5.3}$$

Proof. The first claim is immediate, due to the fact that $\mathbf{m}_t = m - t$ in a simple removal process, and the second then follows from (5.1). □

5.2 Step Decompositions and Incremental Weight

In order to invoke the differential equations method, it is necessary to predict the expected increment a process based on information available at time t. Ideally, then, we would have a set of continuous functions $F_i : \Phi \to \mathbb{R}$ such that

$$\mathsf{E}_t[\Delta\boldsymbol{\mu}_{t+1}(i)] = F_i(\boldsymbol{\mu}_t), \tag{5.4}$$

or the equivalent for the degree. In this situation, the pointwise Lipschitz conditions, along with the equicontinuity of $(\boldsymbol{\mu}_\xi)$, would yield a set of w.e.h.p. differential constraints

$$d\mu_\xi(i) = F_i(\mu_\xi)d\xi. \tag{5.5}$$

The ultimate goal, though, is the set of differential constraints in (5.5), and the predictability of expected increments in (5.4) is not the only way to establish this. In applications, we will take a more indirect route and establish several intermediate constraints before deriving such a complete system of differential equations.

This indirect route will involve expressing the low-level behavior of a removal process in terms of occurrences if individual time-dependent events, which yield information about the endpoint $\mathbf{a}_t$ removed from $\mathbf{A}_{t-1}$ to produce $\mathbf{A}_t$. The topological representation of a time-dependent event was described in §4.4.5 of the previous chapter:

- the *indicator process* satisfies $\mathbf{I}_t[L] = 1$ if L_t occurs and $\mathbf{I}_t[L] = 0$ otherwise;
- the *measure process* has increments $\Delta\boldsymbol{\sigma}_{t+1}[L] = \mathbf{I}_t[L]$, and initial value $\boldsymbol{\sigma}_0[L] = 0$.

The measure process is implicitly scaled by dividing by the duration $T = m$, so we have $\boldsymbol{\sigma}_\xi[L] = \boldsymbol{\sigma}_{\lfloor \xi m \rfloor}[L]/m$.

The reasons for breaking down a process into time-dependent events are twofold. First, there is the issue of algebraic convenience: the combinatorial behavior of a process is more naturally described in terms of discrete events rather than continuous functions. Once we have computed differential constraints for these measure processes, we can then change variables to derive differential constraints for the scaled parameters μ_ξ or λ_ξ.

The second, and perhaps more important reason, is that in most cases, a relationship in (5.4) will not actually hold. This is because the processes we study will often exhibit non-homogeneous behavior, in which the method by which $\mathbf{a}_t$ is chosen from $\mathbf{A}_{t-1}$ may change drastically from one step to the next. In this case, the expected increments cannot be expressed in terms of a single continuous function.

In this section, we address these issues by defining two different kinds of step decompositions:

- in §5.2.1, we define the *step degree*, which is just the (true) degree $\mathbf{deg}_{t-1}(\mathbf{a}_t)$ of the endpoint $\mathbf{a}_t$ at the time it is removed;
- in §5.2.2, we discuss decomposition by *selection method*, which is simply the method used to choose $\mathbf{a}_t$ from $\mathbf{A}_{t-1}$, and unlike the step degree $\mathbf{deg}_{t-1}$, is knowable at time $t-1$.

Then, in §5.2.3, we discuss a specific decomposition based on *random* versus *deterministic* selection methods.

Finally, in §5.2.4, we introduce an additional change of variables in which each step is weighted according to the inverse $1/\mathbf{m}_t$ of the size of the remaining endpoint set $\mathbf{m}_t = |\mathbf{A}_t|$. The idea is that the endpoint $\mathbf{a}_{t+1}$ removed from $\mathbf{A}_t$ represents a fraction of $1/\mathbf{m}_t$ of the remaining endpoints, and for this reason, the weighted time scale often more convenient algebraically.

5.2.1 Step Degrees

In any simple removal process, exactly one endpoint is removed each step, and hence we may define a basic step decomposition as follows.

Definition 5.2.1. The *degree* of a step t is the random variable $\mathbf{deg}_{t-1}(\mathbf{a}_t)$ corresponding to the degree of the endpoint $\mathbf{a}_t$ immediately before $\mathbf{a}_t$ is removed. We define a corresponding step decomposition by

$$\Delta\boldsymbol{\sigma}_t[i] = \begin{cases} 1 \text{ if } \mathbf{deg}_{t-1}(\mathbf{a}_t) = i \\ 0 \text{ otherwise.} \end{cases}$$

These step-processes $\boldsymbol{\sigma}_t[i]$ are not predictable, since the degree of the vertex $\mathbf{a}_t$ is not (generally) determined until time t. Hence, $\Delta\boldsymbol{\sigma}_{t+1}[i]$ cannot be expressed as the indicator for an event which occurs at time t, since this depends on $\mathbf{a}_{t+1}$, which has not yet been selected at time t.

The step degree is thus not directly useful in computing the increments $\Delta\boldsymbol{\lambda}_{t+1}(i)$ or $\Delta\boldsymbol{\mu}_{t+1}(i)$ based on the state at time t. The purpose of this decomposition is simply to describe the relationship between these increments and the degree of the endpoint removed.

Proposition 5.2.1. *For any regular removal process, the following differential constraints hold u.a. for all* $\xi \in [0, 1)$*:*

$$d\lambda_\xi(i) = (d\sigma_\xi[i+1] - d\sigma_\xi[i])M(\lambda), \tag{5.6}$$

$$d\mu_\xi(i) = \frac{(i+1)(d\sigma_\xi[i+2] - d\sigma_\xi[i+1]) + \mu_\xi(i)d\xi}{1-\xi}. \tag{5.7}$$

Proof. Removing an endpoint of degree i changes the degree of a single vertex from i to $i-1$, and since there are $|V| = n$ vertices, we thus have

$$\Delta\boldsymbol{\lambda}_t(i) = \frac{\Delta\boldsymbol{\sigma}_t[i+1] - \Delta\boldsymbol{\sigma}_t[i]}{n}.$$

All step-processes are implicitly scaled by dividing by the duration m, which is equal to $m = M(\boldsymbol{\lambda}_0) \cdot n$, so $m/n \to M(\lambda)$, which yields equation (5.6).

We may now derive (5.7) algebraically, using equation (5.3) from corollary 5.1.2:

$$\begin{aligned} d\mu_\xi(i) = d\left(\frac{(i+1)\lambda_\xi(i+1)}{M(\lambda)(1-\xi)}\right) &= \frac{(i+1)d\lambda_\xi(i+1)}{M(\lambda)(1-\xi)} - \frac{(i+1)\lambda_\xi(i+1)d(1-\xi)}{M(\lambda)(1-\xi)^2} \\ &= \frac{(i+1)(d\sigma_\xi[i+2] - d\sigma_\xi[i+1]) + \mu_\xi(i)d\xi}{1-\xi}. \end{aligned}$$

□

5.2.2 Decomposition by Selection Method

In order to invoke the differential equations method, it is necessary to predict the expected increment a process based on information available at time t. To an extent, the step decomposition defined above simplifies this problem algebraically, since at this point is suffices to predict the expected increments of the step degree processes. This can be accomplished, for example, by establishing a continuous relationship of the form

$$\mathsf{E}_t[\Delta\boldsymbol{\sigma}_t[i]] = F_i(\boldsymbol{\mu}_t).$$

On the other hand, step degrees do not get us any closer to actually determining these increments, since $\mathbf{deg}_t(\mathbf{a}_{t+1})$ is not knowable until time $t+1$. We now discuss an additional step decomposition which will serve an more substantive role.

This decomposition is based on the *selection method*, which is simply the method by which the endpoint $\mathbf{a}_{t+1}$ is chosen from $\mathbf{A}_t$. So as not to introduce unnecessary complications, we will not offer a formal definition of a selection method. What is important is that the method used to select $\mathbf{a}_{t+1}$ is determined at time t. At a technical level, then, for any

selection method L, the indicator random variable $\mathbf{I}_t[L]$ describes the method by which the *next* endpoint $\mathbf{a}_{t+1}$ will be selected.

The exact selection methods in play for any particular application will of course depend on the algorithm we are analyzing, and it is simplest to just define these selection methods, and ensure that the behave in the way we intend, on a case by case basis. For now, we illustrate by listing some common examples:

- choose $\mathbf{a}_{t+1}$ uniformly at random from $\mathbf{A}_t$;
- choose an endpoint $\mathbf{a}_{t+1}$ with minimum degree;
- choose an endpoint $\mathbf{a}_{t+1}$ with degree i for some fixed i (if possible);
- choose an endpoint $\mathbf{a}_{t+1}$ from the same vertex as the previous endpoint $\mathbf{a}_t$ (again, if possible).

If selection methods are chosen properly, then the increments of the step degree processes $\boldsymbol{\sigma}_t[i]$ should be predictable, in expectation, based on the selection method and either the degree or residual distribution at time t. Hence, we would ideally have

$$\mathsf{E}_t[\Delta\boldsymbol{\sigma}_{t+1}[i] \mid \mathbf{I}_t[L] = 1] = F_{L,i}(\boldsymbol{\mu}_t) \tag{5.8}$$

where $F_{L,i}$ is a continuous function mapping $\Phi \to [0,1]$. There are no hard and fast rules, though, and the "validity" of a selection method is measured by whether or not it helps us compute a solution, rather than by any technical criteria.

Decomposition by Selection Method and Degree

In order to exploit the condition in (5.8), we will ultimately decompose steps by both selection method and degree by defining

$$\Delta\boldsymbol{\sigma}_{t+1}[L(i)] = \mathbf{I}_t[L] \cdot \Delta\boldsymbol{\sigma}_{t+1}[i] \tag{5.9}$$

Since $\mathbf{I}_t[L]$ indicates the method by which $\mathbf{a}_{t+1}$ is selected, then $\boldsymbol{\sigma}_t[L(i)]$ counts the total number of times an endpoint of degree i was selected by the method L, during the interval $1 \leq s \leq t$.

This process is clearly Lipschitz, and is therefore w.e.h.p. determined by the sum of its expected, rather than true increments. Hence, if (5.8) holds, then we have a w.e.h.p. differential constraint

$$d\sigma_\xi[L(i)] = F_{L,i}(\mu_\xi) d\sigma_\xi[L].$$

Moreover, if a particular algorithm only uses finitely many selection methods $\{L_j\}_{j\in J}$, and if each of these selection methods satisfy a relationship of the form (5.9), then we will have a w.e.h.p. differential constraint

$$d\sigma_\xi[i] = \sum_{j\in J} F_{j,i}(\mu_\xi) d\sigma_\xi[L_j]$$

which allows us to determine the total frequency of steps of each degree i.

This does not yet allow us to solve for μ_ξ, since we still need to compute solutions to each $d\sigma_\xi[L_j]$, and thus another set of differential constraints is required. The way to do this will vary depending on the application, so we will not go into details at this time.

5.2.3 Random and Deterministic Selection

While the exact selection methods used in any given removal process will vary depending on the algorithm we are analyzing, one basic, and fairly universal, decomposition is into *random* and *deterministic* steps. As the name suggests, in a random step, the endpoint $\mathbf{a}_{t+1}$ is chosen uniformly at random from $\mathbf{A}_t$. Deterministic selection is "anything else," and our basic notation is:

- $\mathbf{I}_t[R]$ indicates uniformly random selection at time $t+1$;
- $\mathbf{I}_t[D]$ indicates deterministic selection at time $t+1$;

Now, in the basic configuration model (CM) algorithm, all even-numbered endpoints must be chosen uniformly at random. Since R_t determines the selection method of the *next* endpoint, then in the basic CM algorithm, we have $\mathbf{I}_t[R] = 1$ whenever t is *odd*, and hence $\Delta\boldsymbol{\sigma}_t[R] = 1$ whenever t is *even*.

Of course, if we are not using the basic CM algorithm, then it will not necessarily be the case that all even selections are uniformly random. Also, even in the basic CM algorithm it is not necessarily the case that all odd selections will be deterministic, since we might choose to select some odd-numbered endpoints uniformly at random, even though we have the ability to specify a deterministic selection method.

In any case, all of the algorithms we analyze will involve random selection in one form or another. Random selection will be discussed in detail in the next section; for now we simply derive the expected increments of step degrees for random selection as follows.

Proposition 5.2.2. *For any regular removal process, the differential constraint*

$$d\sigma_\xi[R(i)] = \mu_\xi(i-1)d\sigma_\xi[R]$$

holds w.e.h.p. for every i and every $\xi \in [0,1)$

Proof. By definition, the residual distribution is the distribution of the residual degree of a uniformly random endpoint, and since true degree is one greater than residual degree, then

$$\mathsf{P}_t[\mathbf{deg}_t(\mathbf{a}_t) = i \mid \mathbf{I}_t[R] = 1] = \boldsymbol{\mu}_t(i-1).$$

It follows that

$$\mathsf{E}_t[\Delta\boldsymbol{\sigma}_{t+1}[R(i)]] = \boldsymbol{\mu}_t(i-1)\Delta\boldsymbol{\sigma}_{t+1}[R],$$

and since $(\boldsymbol{\sigma}_t[R(i)])$ is a Lipschitz process, then by martingale concentration it suffices to sum over expected rather than true increments. Finally, since both $(\boldsymbol{\mu}_t(i-1))$ and $(\sigma_t[R])$ are well-scaled over any interval $t/m \in [0, 1-\epsilon]$, the above differential constraint follows from convergence of the Riemann-Stieltjes integral in proposition 4.4.3. □

Using the partial derivative notation of §4.4.5 from the previous chapter, we may express this differential constraint as

$$\frac{\partial\sigma_\xi[i]}{\partial\sigma_\xi[R]} = \mu_\xi(i-1).$$

As explained in §4.4.5, this is not really a "partial derivative," but just a notational abbreviation to describe the change on $\sigma_\xi[i]$ which is due to random selections.

We can also express the change in both the degree and residual distribution due to random selections as follows.

Corollary 5.2.3. *For any regular removal process, the following hold w.e.h.p. for $\xi \in [0,1)$:*

$$\frac{\partial \mu_\xi(i)}{\partial \sigma_\xi[R]} = \frac{(i+1)\mu_\xi(i+1) - i\mu_\xi(i)}{1-\xi}, \tag{5.10}$$

$$\frac{\partial \lambda_\xi(i)}{\partial \sigma_\xi[R]} = \frac{(i+1)\lambda_\xi(i+1) - i\lambda_\xi(i)}{1-\xi}. \tag{5.11}$$

Proof. By proposition 5.2.1, the increments of the residual distribution can be expressed in terms of step degrees by

$$d\mu_\xi(i) = \frac{(i+1)(d\sigma_\xi[i+2] - d\sigma_\xi[i+1]) + \mu_\xi(i)d\xi}{1-\xi},$$

and combining this equation with $d\sigma_\xi[R(i)] = \mu_\xi(i-1)d\sigma_\xi[R]$ yields

$$\begin{aligned}\frac{\partial \mu_\xi(i)}{\partial \sigma_\xi[R]} &= \frac{(i+1)(\mu(i+1) - \mu(i)) + \mu_\xi(i)}{1-\xi} \\ &= \frac{(i+1)\mu(i+1) - i\mu(i)}{1-\xi}.\end{aligned}$$

For the degree distribution, again by proposition 5.2.1, we have $d\lambda_\xi(i) = (d\sigma_\xi[i+1] - d\sigma_\xi[i])M(\lambda)$. We also recall that $\mu_\xi(i) = \frac{(i+1)\lambda_\xi(i+1)}{M(\lambda)(1-\xi)}$, so once again

$$\begin{aligned}\frac{\partial \lambda_\xi(i)}{\partial \sigma_\xi[R]} &= \frac{\mu_\xi(i) - \mu_\xi(i-1)}{M(\lambda)} \\ &= \frac{(i+1)\lambda_\xi(i+1) - i\lambda_\xi(i)}{1-\xi}.\end{aligned}$$

□

5.2.4 Incremental Weighting

Unlike the degree distribution, the increments of the residual distribution, as computed in (5.7), depend on the time as well as the degree of the endpoint removed at time t.

Moreover, for uniformly random selection, as shown above, the derivatives of both the degree and residual distributions are non-homogeneous in that they depend on ξ.

Intuitively, this is because a single endpoint constitutes a larger fraction of the $\mathbf{A}_t$ for larger values of t, and therefore the effect of removing $\mathbf{a}_{t+1}$ is magnified accordingly. In order to achieve time-homogeneity with respect to the residual distribution, we will often re-parametrize the time scale as follows by defining the *incremental weight* of a given step to be

$$\Delta \mathbf{w}_{t+1} = 1/\mathbf{m}_t. \tag{5.12}$$

where $\mathbf{m}_t = |\mathbf{A}_t|$.

The idea is that the endpoint $\mathbf{a}_{t+1}$ represents a fraction of $\mathbf{w}_{t+1}$ of the remaining endpoint set $\mathbf{A}_t$ from which it is removed. We shall use incremental weight as a change of measure, so for any event L, we define the weighted measure process by

$$\Delta \mathbf{w}_{t+1}[L] = \Delta \boldsymbol{\sigma}_{t+1}[L] \cdot \Delta \mathbf{w}_{t+1} = \Delta \boldsymbol{\sigma}_{t+1}/\mathbf{m}_t.$$

Since the increments of $\mathbf{w}_t$ already have size $O(1/m)$ (provided that $t < (1-\epsilon)m$, then this process does not need scaling. Also, since $\mathbf{m}_t = m - t$, then the process $\mathbf{w}_t$ is not random, and there is only one solution, which is given by the differential equation

$$dw_\xi = \frac{d\xi}{1-\xi}.$$

The solution to this differential equation is $w_\xi = -\ln(1-\xi)$, or conversely $\xi = 1 - e^{-w_\xi}$.

Note that the weighted time scale can only be used if we restrict our attention to a time interval $\xi \in [0, 1-\epsilon]$, due to the fact that $\lim_{\xi \to 1} w_\xi = \infty$. The weighted time scale is primarily useful in conjunction with the residual distribution, which as we recall can also only be used for an interval $\xi = [0, 1-\epsilon]$. Moreover, by making the constant $\epsilon > 0$ arbitrarily small, we may effectively compute a solution for all $\xi \in [0, 1)$ using this change of variables.

For any time-dependent event L, the solution to the weighed measure process $\mathbf{w}_t[L]$ will satisfy $dw_\xi[L] = \frac{d\sigma_\xi[L]}{1-\xi}$. Accordingly, the relationship between step degrees and the

residual distribution can be expressed by

$$d\mu_\xi(i) = (i+1)(dw_\xi[i+2] - dw_\xi[i+1]) + \mu_\xi(i)dw_\xi. \tag{5.13}$$

Intuitively, the "extra term" $\mu_\xi(i)dw_\xi$ accounts for the fact, whenever any endpoint is removed from $\mathbf{A}_t$, of any degree, each remaining endpoint now constitutes a slightly larger fraction of the endpoint set $\mathbf{A}_{t+1}$.

The differential constraints for random selection from corollary 5.2.3 can also be stated in terms of incremental weight by

$$\frac{d\mu_\xi(i)}{\partial w_\xi[R]} = (i+1)\mu_\xi(i+1) + i\mu_\xi(i). \tag{5.14}$$

Also, since differential constraints for the degree distribution with respect to random selection are the same as for the residual distribution, then (5.14) will hold identically for λ_ξ as well.

5.3 Random Endpoint Removal

We now examine the effect of random endpoint removal in more detail. We begin by solving the system of differential equations

$$d\mu_\xi(i) = \Big((i+1)\mu_\xi(i+1) + i\mu_\xi(i)\Big)dw_\xi \tag{5.15}$$

which correspond to random selection as discussed above.

The solution, which is computed in §5.3.1 is only directly applicable to a process in which every endpoint is removed uniformly at random. In other cases, we must consider the combined effect of random and deterministic selections. However, as shown in §5.3.2, the solution to (5.15) is also helpful when deterministic selection is also involved.

5.3.1 Unique Solution to (5.15)

We now solve the above system of differential equations. For simplicity, we assume that the initial state is $\xi = 0$, but this is clearly without loss of generality, and the following holds for any interval $[\zeta_0, \zeta_1]$ by changing variables appropriately.

Theorem 5.3.1. *For any $\mu \in \mathbb{R}^\omega$ with $\|\mu\|_1 < \infty$, any interval $[0, \zeta_1] \subseteq [0, 1)$, any continuous, nondecreasing function (w_ξ) on $\xi \in [0, \zeta_1]$, with $w_0 = 0$, we define*

$$\mu_\xi(i) = \sum_{j=i}^{\infty} \mu(j) \binom{j}{i} (1 - e^{-w_\xi})^{j-i} e^{-iw_\xi}. \tag{5.16}$$

for $\xi \in [0, \zeta_1]$. Then:

1. *this function satisfies (5.15) for all $\xi \in [0, \zeta_1)$, with initial condition $\mu_0 = \mu$,*
2. *$\|\mu_\xi\|_1 < \infty$ for all $\xi \in [0, \zeta_1]$ and the function (μ_ξ) is continuous in ℓ_1,*
3. *(μ_ξ) uniquely satisfies these two conditions,*
4. *the same is true if the coordinate set is restricted to $i \in \{j, j+1, \ldots\}$ for some $j > 0$.*

We briefly note that (5.15) is being treated as just an infinite system of differential equations in $\mathbb{R}$, and not as a single differential equation in ℓ_1. The bound on $\|\mu_\xi\|_1$ serves only to rule out potential non-analytic solutions.

Proof. We first verify that this solution is correct:

$$\begin{aligned}
\frac{d\mu_\xi(i)}{dw_\xi} &= \sum_{j=i}^{\infty} \mu(j) \binom{j}{i} \Big((j-i) e^{-w_\xi} (1 - e^{-w_\xi})^{j-i-1} e^{-iw_\xi} - i(1 - e^{-w_\xi})^{j-i} e^{-iw_\xi} \Big) \\
&= (i+1) \left(\sum_{j=i+1}^{\infty} \mu(j) \binom{j}{i+1} (1 - e^{-w_\xi})^{j-(i+1)} e^{-(i+1)w_\xi} \right) - i\mu_\xi(i) \\
&= (i+1)\mu_\xi(i+1) - i\mu_\xi(i).
\end{aligned}$$

It is evident that $\mu_0 = \mu$, so the first claim is proved.

The bound on $\|\mu_\xi\|_1$ can be deduced by defining

$$\mu_\xi(i, j) = \mu(j) \binom{j}{i} (1 - e^{-w_\xi})^{j-i} e^{-iw_\xi}$$

so $\mu_\xi(i) = \sum_{j=0}^{\infty} \mu_\xi(i, j)$. Each function $\mu_\xi(\cdot, j)$ is finite-dimensional and therefore continuous in ℓ_1. Also, we have $\|\mu_\xi(\cdot, j)\|_1 = |\mu(j)|$, and therefore

$$\|\mu_\xi\|_1 \leq \sum_{j=0}^{\infty} \|\mu_\xi(\cdot, j)\|_1 = \|\mu\|_1 .$$

Moreover, due to the fact that $\|\mu\|_1 < \infty$, the partial sums $\sum_{j \leq J}(\mu_\xi(\cdot, j))$ converge uniformly to (μ_ξ) in ℓ_1 as $J \to \infty$, hence (μ_ξ) is continuous in ℓ_1.

For uniqueness, note that (5.15) is linear, so if we have two solutions (μ_ξ) and (μ'_ξ), then $(\nu_\xi) = (\mu_\xi) - (\mu'_\xi)$ is also a solution, with initial state $\nu_0(i) = 0$ for all i. This function (ν_ξ) will also be bounded and continuous in ℓ_1 and we now show that such a function cannot exist.

It is clear that for any j, given the initial condition $\nu_\xi(i) = 0$ for all i, and the value of $\nu_\xi(j+1)$ for all $\xi \in [0, 1)$, there is a unique solution to the finite dimensional system consisting of the $\nu_\xi(i)$ for $0 \leq i \leq j+1$. This solution will be given by

$$\nu_\xi(i) = \int_{\zeta=0}^{\zeta_1} (j+1)\binom{j}{i}\nu_\zeta(j+1)(1 - e^{w_\zeta - w_\xi})^{j-i} e^{i(w_\zeta - w_\xi)} dw_\zeta, \tag{5.17}$$

which can be verified by performing the same calculation as above, but inside the integral.

Let us now define

$$f(i) = \sup\{|\nu_\xi(i)| : \xi \in [0, \zeta_1]\},$$

and (5.17) yields

$$\begin{aligned} f(i) &\leq \int_{\zeta=0}^{\zeta_1} (j+1)\binom{j}{i} |\nu_\zeta(j+1)| \, (1 - e^{w_\zeta - w_{\zeta_1}})^{j-i} e^{i(w_\zeta - w_\xi)} dw_\zeta \\ &\leq (j+1)\binom{j}{i} f(j+1)(1 - e^{w_{\zeta_1}})^{j-i} w_{\zeta_1}. \end{aligned}$$

By changing variables $h = j + 1$ we may deduce

$$f(h) \geq f(i) h^{-i} \epsilon^{-h} C(i),$$

where $\epsilon = (1 - e^{-w_{\zeta_1}}) < 1$, and $C(i) = \epsilon^{i+1}/w_{\zeta_1} > 0$. In this case, if $f(i) > 0$ for any given i, then we must have $\lim_{h \to \infty} f(h) = \infty$, which would imply that $\sup\{\|\nu_\xi\|_1 : \xi \in [0, \zeta_1]\} = \infty$. Hence, any solution to (5.15), other than the function (μ_ξ) in (5.16) does not belong to ℓ_1.

The final claim, that we may restrict the system of differential equations to coordinates $i \in \{j, j+1, \ldots\}$ is immediate, since the expression for $\mu_\xi(i)$ in (5.16) only depends on $j \geq i$. The uniqueness of such a solution (in ℓ_1) follows by an identical argument. □

This theorem can be used to compute a unique solution to an endpoint removal process consisting entirely of random selections as follows.

Corollary 5.3.2. *For any endpoint removal process in which every step is random, the unique w.e.h.p. solutions to the degree and residual distribution are given by*

$$\mu_\xi(i) = \sum_{j=i}^{\infty} \mu(j) \binom{j}{i} \xi^{j-i}(1-\xi)^i \quad \textit{and} \quad \lambda_\xi(i) = \sum_{j=i}^{\infty} \lambda(j) \binom{j}{i} \xi^{j-i}(1-\xi)^i$$

for all $\xi \in [0, 1)$.

Proof. If every step is random, then $w_\xi[R] = w_\xi = \ln(1-\xi)$, so $e^{-\xi} = 1 - \xi$. By theorem 5.3.1 proved above, the above solutions are unique in ℓ_1 for any interval $[0, 1-\epsilon]$, which implies uniqueness over all $[0, 1)$. □

As discussed in section 5.1, solutions to the residual distribution cannot be extended to $\xi = 1$, so for μ_ξ the above solution only holds on $\xi \in [0, 1)$, despite the fact that, algebraically, the solution can be extended to $\xi = 1$ by continuity. For λ_ξ, the solution does indeed holds for $\xi = 1$, but this is somewhat trivial, since $\lambda_1(0) = 1$ and $\lambda_1(i) = 0$ for $i > 0$, so all this tells us is that there are no more endpoints left at the end of the process.

5.3.2 Combined Random and Deterministic Selection

For general removal processes, not all steps will involve random selection, and hence the solutions computed above are not immediately applicable. In many cases, though, we may use the solution to random selection as a starting point to construct a solution to a general process which include both kinds of selections.

This is because deterministic selections are usually chosen from vertices of low degree, and therefore the solution in (5.16) can be used to take care of the upper tail of the distribution. What remains will be a finite system of differential equations, which is generally be easier to solve, and does not present any technical difficulties related to uniqueness. Again, we state this corollary for an interval $[0, \zeta_1)$, but this is without loss of generality

with respect to the starting point, and holds just as well for $[\zeta_0, \zeta_1)$ modulo the appropriate change of variables.

Corollary 5.3.3. *Assume a regular removal process satisfies $d\sigma_\xi[D(i)] = 0$ for all $i \geq k$ all $\xi \in [0, \zeta_1)$. Then:*

1. *for any $i \geq k$ and any $\xi \in [0, \zeta_1)$, every w.e.h.p. solution to λ_ξ satisfies*

$$\lambda_\xi(i) = \sum_{j=i}^{\infty} \lambda(j) \binom{j}{i} (1 - e^{-w_\xi[R]})^{j-i} e^{-i w_\xi[R]}, \tag{5.18}$$

2. *for any $i \geq k-1$ and any $\xi \in [0, \zeta_1)$, every w.e.h.p. solution to μ_ξ satisfies*

$$\mu_\xi(i) = e^{w_\xi[D]} \sum_{j=i}^{\infty} \mu(j) \binom{j}{i} (1 - e^{-w_\xi[R]})^{j-i} e^{-i w_\xi[R]}. \tag{5.19}$$

Proof. For the degree distribution, the above is immediate, since if $d\sigma_\xi[D(i)] = 0$ for $i \geq k$, then we simply have $d\lambda_\xi(i) = \big((i+1)\lambda_\xi(i+1) - i\lambda(i)\big) dw_\xi[R]$, which was shown above to have a unique solution.

For the residual distribution, since residual degree is one less than true degree, the above will hold for $\mu_\xi(i)$ where $i \geq k-1$. One possible way to compute (5.19) is to use the algebraic relationship between λ_ξ and μ_ξ, along with the fact that $w_\xi[R] + w_\xi[D] = w_\xi = \ln(1-\xi)$.

A quicker method is to note that, since $dw_\xi[D(i+1)] = 0$, we have the differential constraint

$$d\mu_\xi(i) = \Big((i+1)\mu_\xi(i+1) - \mu_\xi(i)\Big) dw_\xi[R] + \mu_\xi(i) dw_\xi[D]. \tag{5.20}$$

We may now consider (5.19) as a function of two variables $w_\xi[R]$ and $w_\xi[D]$, with partial derivatives

$$\frac{\partial \mu_\xi(i)}{\partial w_\xi[R]} = (i+1)\mu_\xi(i+1) - i\mu_\xi(i),$$
$$\frac{\partial \mu_\xi(i)}{\partial w_\xi[D]} = \mu_\xi(i),$$

which yields (5.20). ∎

5.4 The Probability Generating Function

For any distribution μ on the set $\mathbb{Z}^*$ of non-negative integers, the *probability generating function* is the power series

$$\psi_\mu(z) = \sum_{i=0}^{\infty} \mu(i) z^i.$$

The p.g.f. is a classical tool in probability theory (e.g. [42]) and it plays a central role in the theory of branching processes [4]. We shall make extensive use of the p.g.f. to facilitate algebraic manipulations of the degree and residual distributions in our analysis of endpoint removal processes.

5.4.1 Basic Properties

We begin by discussing some basic properties of the p.g.f. For any distribution μ on $\mathbb{Z}^*$, the power series $\psi_\mu(z)$ is absolutely convergent for $z \in [-1, 1]$. Moreover:

- $\psi_\mu(1) = \sum_i 1^i \mu(i) = 1$ and $\psi_\mu(0) = \sum_i 0^i \mu(i) = \mu(0)$,
- for any $z \in [0, 1)$ all of the derivatives of $\psi_\mu(z)$ are non-negative.

Accordingly, we will typically think of ψ_μ as a monotonic function mapping the closed interval $[0, 1]$ to itself, and such that $\psi_\mu(1) = 1$.

The derivatives of ψ_μ at the values 0 and 1 carry special significance, since they yield probability weights and factorial moments:

$$\frac{\psi_\mu^{(i)}(0)}{i!} = \mu(i)$$
$$\psi_\mu^{(i)}(1) = \sum_{j=i}^{\infty} (j)_i \mu(j) = M_i(\mu).$$

It is often useful to think of the p.g.f. "probabilistically" in the sense that, if $\mathbf{x}_\mu$ denotes a μ-distributed random variable, then we have

$$\psi_\mu(z) = \mathsf{E}[z^{\mathbf{x}_\mu}].$$

As such, the p.g.f. closely resembles the moment-generating function $f(t) = \mathsf{E}[e^{t\mathbf{x}_\mu}]$, since the two are trivially related by $f(t) = \psi_\mu(e^t)$. The p.g.f. thus exhibits the same general kinds of properties as the moment-generating function or the characteristic function; for instance, given independent random variables $\mathbf{x}_\nu, \mathbf{x}_\mu$, we have

$$\mathsf{E}[z^{\mathbf{x}_\nu + \mathbf{x}_\mu}] = \mathsf{E}[z^{\mathbf{x}_\nu}]\mathsf{E}[z^{\mathbf{x}_\mu}] = \psi_\mu(z)\psi_\nu(z),$$

and thus summation of independent random variables corresponds to multiplication of generating functions.

However, there are certain properties of the p.g.f. which are uniquely applicable to integer-valued random variables, and for this reason, the p.g.f. is often more powerful than the moment-generating function for distributions on $\mathbb{Z}^*$. A particularly relevant example is the following. Assume that for distributions μ and ν, we define a random variable $\mathbf{y}$ by summing a random, μ-distributed number of i.i.d. copies of $\mathbf{x}_\nu$ as follows:

$$\mathbf{y} = \sum_{i=1}^{\mathbf{x}_\mu} \mathbf{x}_{\nu,i}.$$

In this case, we may compute

$$\mathsf{E}[z^{\mathbf{y}}] = \textstyle\sum_j \mu(j)\mathsf{E}\left[\prod_{i=1}^{j} z^{\mathbf{x}_{\nu,i}}\right] = \sum_j \mu(j)\psi_\nu(z)^j = \psi_\mu(\psi_\nu(z)).$$

Due in part to the relationship between composition of the p.g.f. and the sum of a random number of random variables, the p.g.f. plays an important role in the theory of branching processes. Branching processes are in turn closely related to random graphs; this connection will be explored in chapters 7 and 8.

5.4.2 The Fixed Point

An important characteristic of a p.g.f. is the location of the smallest fixed point in the interval $[0, 1]$, which we will generally denote by z_μ. Some properties of this fixed point are described in the following proposition.

Proposition 5.4.1. *For any distribution μ:*

1. *if* $M(\mu) \leq 1$ *and* $\mu(1) < 1$ *then* $z_\mu = 1$*;*

2. *if* $M(\mu) > 1$*, then* $z_\mu < 1$*, the derivative at* z_μ *satisfies* $\psi'_\mu(z_\mu) < 1$*, and* z_μ *is the unique fixed point in* $[0, 1)$*;*

3. *if* $\mu(0) = 0$ *then* $z_\mu = 0$*.*

Proof. For the first claim, recall that $\psi_\mu(1) = 1$ and $M(\mu) = \psi'_\mu(1)$, and since ψ_μ has non-decreasing derivatives, then if $M(\mu) \leq 1$, we must have $\psi_\mu(z) \geq z$ for all $z \in [0, 1]$. Moreover, unless $\psi_\mu(z) = z$ for all z, then it must in fact be the case that $\psi_\mu(z) > z$ for all $z \in [0, 1)$, in which case $z_\mu = 1$. The case $\psi_\mu(z) = z$ for all z occurs if and only if $\mu(1) = 1$ and thus $\mu(i) = 0$ for all $i \neq 1$.

For the second claim, if $M(\mu) > 1$, then $\psi_\mu(1 - \epsilon) < 1 - \epsilon$ for ϵ sufficiently small, but since $\psi(0) \geq 0$, then there must be a fixed point $z_\mu \in [0, 1)$. Also, it is evident that we must have $\psi'_\mu(z_\mu) < 1$ at this fixed point, and hence, as reasoned above, it must be the case that $\psi_\mu(z) > \psi_\mu(z)$ for $z \in [0, z_\mu)$.

Finally, since $\psi_\mu(0) = \mu(0)$, then clearly if $\mu(0) = 0$ then $z_\mu = 0$. □

In particular, note that, unless $\mu(1) = 1$, the question of whether the fixed point z_μ is strictly less than 1 depends only on $M(\mu)$. Accordingly, we shall categorize distributions as follows

- μ is *super-critical* if $M(\mu) > 1$ (and hence $z_\mu < 1$);
- μ is *sub-critical* if $M(\mu) < 1$ (and hence $z_\mu = 1$);
- μ is *critical* if $M(\mu) = 1$.

If μ is critical, then $z_\mu = 1$ if and only if $\mu(1) = 1$. Intuitively, in this case the degree distribution will satisfy $\lambda(2) = 1 - \lambda(0)$, and hence all but $o(n)$ non-empty vertices in the corresponding random graph will have degree 2. We will typically ignore this situation, in which case we have the following corollary.

Corollary 5.4.2. *For any distribution with $\mu(1) < 1$, we have*

$$z_\mu = \inf\{z \in [0,1] : \psi_\mu(z) < z\} = \sup\{z \in [0,1] : \psi_\mu(z) > z\},$$

and hence, for any $z \in [0,1)$:

1. *$\psi(z) > z$ if and only if $z < z_\mu$;*
2. *$\psi(z) < z$ if and only if $z > z_\mu$.*

Proof. Immediate. □

If $\mu(1) = 1$, then the the location of z_μ is unstable with respect to small changes in μ since, for example, any distribution with $\mu(1) = 1-\epsilon$ and $\mu(0) = \epsilon$, will have $M(\mu) < 1$ and thus $z_\mu = 1$. Accordingly, in this situation, the location of z_μ is not useful for asymptotic analysis, since our general assumptions only guarantee that the true residual distribution belongs to an arbitrarily small neighborhood of the limiting distribution μ.

Plotting the P.G.F.

A useful way to understand the probability generating function at a qualitative level is to visually examine its plot in the unit square $[0,1] \times [0,1]$. In figure 5.1, we have plotted some typical generating functions to illustrate the various criticality conditions discussed above in relation to the fixed point z_μ. In particular, note that, for values of $0 < z < 1$, larger value of $\psi_\mu(z) = \mathsf{E}[x^{\mathbf{x}_\mu}]$ corresponds to a "smaller" random variable $\mathbf{x}_\mu$.

5.4.3 The Degree and Residual Distribution

In the context of a random graph $\boldsymbol{\mathcal{G}}(\mu)$, we will generally work the p.g.f. of the residual distribution rather than the degree distribution. However, ψ_μ can be computed easily from ψ_λ from the formula $\mu(i) = \frac{(i+1)\lambda(i+1)}{M(\lambda)}$ as follows:

$$\begin{aligned}\psi_\mu(z) = \sum_{i=0}^{\infty} \mu(i) z^i &= \sum_{i=0}^{\infty} \frac{(i+1)\lambda(i+1)z^i}{M(\lambda)} \\ &= \frac{\psi'_\lambda(z)}{\psi'_\lambda(1)}.\end{aligned}$$

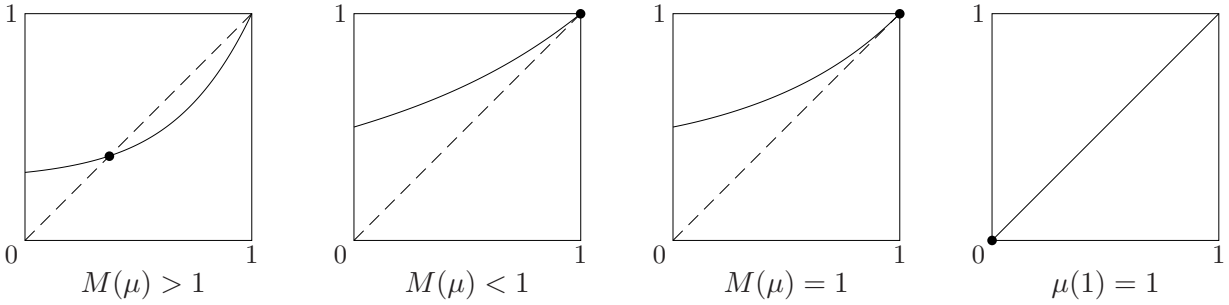


Figure 5.1: Typical p.g.f. plots, including the fixed point z_μ and the dashed line $f(z) = z$. From the left, the super-critical case $M(\mu) > 1$, the sub-critical case $M(\mu) < 1$, the critical case $M(\mu) = 1$, and the special case $\mu(1) = 1$.

The p.g.f. ψ_λ cannot be recovered from ψ_μ, since $\lambda(0)$ cannot be determined from the residual distribution. However, we may compute the formal integral

$$\int \psi_\mu(z) = \sum_{i=0}^{\infty} \frac{\mu(i) z^{i+1}}{i+1} = \sum_{i=0}^{\infty} \frac{\lambda(i+1) z^{i+1}}{M(\lambda)} = M(\lambda)(\psi_\lambda(z) - \lambda(0)),$$

and since $\lambda(0) = \psi_\lambda(0)$, we have

$$\psi_\lambda(z) = \lambda(0) + (1 - \lambda(0)) \left(\frac{\int_{y=0}^{z} \psi_\mu(y) dy}{\int_{y=0}^{1} \psi_\mu(y) dy} \right).$$

Since $\lambda(0)$ corresponds to the fraction of vertices which have degree 0, for practical purposes, we can usually simply assume that $\lambda(0) = 0$. In this case, the above equation yields a useful expression for the first moment of the degree distribution:

$$M(\lambda) = \left(\int_{z=0}^{1} \psi_\mu(z) \right)^{-1}. \tag{5.21}$$

5.4.4 The P.G.F. and Random Endpoint Selection

In section 5.3, we computed the solution for both the degree and residual distributions for a removal process in which a single endpoint is removed uniformly at random at each time step. The solution was the same in both cases:

$$\mu_\xi(i) = \sum_{j=i}^{\infty} \mu(j) \binom{j}{i} \xi^{j-i} (1 - \xi)^i. \tag{5.22}$$

While this formula is not overly burdensome, the distribution μ_ξ can be described much more succinctly in terms of its generating function.

Proposition 5.4.3. *The generating function for the distribution μ_ξ in* (5.22) *is*

$$\psi_\xi(z) = \psi_\mu(\xi + (1-\xi)z). \tag{5.23}$$

Proof. We compute:

$$\begin{aligned}
\psi_\xi(z) = \sum_{i=0}^{\infty} \mu_\xi(i) z^i &= \sum_{i=0}^{\infty} \sum_{j=i}^{\infty} \mu(j) \binom{j}{i} \xi^{j-i} (1-\xi)^i z^i \\
&= \sum_{j=0}^{\infty} \mu(j) \sum_{i=0}^{j} \binom{j}{i} \xi^{j-i} (1-\xi)^i z^i \\
&= \sum_{j=0}^{\infty} \mu(j) (\xi + (1-\xi)z)^j \\
&= \psi_\mu(\xi + (1-\xi)z).
\end{aligned}$$

□

This simple expression for $\psi_\xi(z)$ can also be derived in a more intuitive, probabilistic manner. The distribution $\mu_\xi(i)$ is in fact a mixture of binomial distributions, and hence a μ_ξ-distributed random variable $\mathbf{x}_{\mu_\xi}$ can be constructed by first generating a random variable $\mathbf{x}_\mu$, and setting $\mathbf{x}_{\mu_\xi}$ equal to a binomial distributed random variable with parameters $\mathbf{x}_\mu$ and $(1-\xi)$. A binomial random variable can in turn be expressed as the sum of independent Bernoulli random variables (i.e. $\{0,1\}$-valued random variables), so we may express $\mathbf{x}_{\mu_\xi}$ by

$$\mathbf{x}_{\mu_\xi} = \sum_{i=1}^{\mathbf{x}_\mu} \mathbf{y}_i$$

where each $\mathbf{y}_i$ is an independent Bernoulli random variable with $\mathsf{E}[\mathbf{y}_i] = \mathsf{P}[\mathbf{y}_i = 1] = 1 - \xi$.

The p.g.f. for a Bernoulli distribution with expected value p is simply $\psi(z) = (1 - p) + pz$ and thus the function ψ_ξ can be computed, as described in §5.4.1, by composition of generating functions, in which case we once again have

$$\psi_\xi = \psi_\mu(\xi + (1-\xi)z).$$

Before proceeding, we make two final observations about this generating function. First, recall that the degree and residual distributions have equivalent solutions for a process in which every selection is made uniformly at random. This fact can also be ascertained easily via generating functions, using the equation

$$\psi_\mu(z) = \frac{\psi'_\lambda(z)}{\psi'_\lambda(1)}.$$

Since $\psi_{\lambda_\xi}(z) = \psi_\lambda(\xi+(1-\xi)z)$, we thus have $\psi'_{\lambda_\xi}(z) = (1-\xi)\psi'_\lambda(\xi+(1-\xi)z)$, and therefore

$$\begin{aligned}\psi_{\mu_\xi}(z) &= \frac{(1-\xi)\psi'_\lambda(\xi+(1-\xi)z)}{(1-\xi)\psi'_\lambda(1)} \\ &= \psi_\mu(\xi+(1-\xi)z).\end{aligned}$$

Finally, not only can the solution to random endpoint removal be expressed in terms of generation functions, but the system of differential equations can be computed and solved using the p.g.f. as well. Recall that, during random endpoint removal, we have $\frac{d}{dw_\xi}\mu_\xi(i) = (i+1)\mu_\xi(i+1) - i\mu_\xi(i)$, and hence we may compute

$$\begin{aligned}\frac{d}{dw_\xi}\psi_\xi(z) &= \sum_{i=0}^{\infty}((i+1)\mu(i+1) - i\mu(i))z^i \\ &= \left(\sum_{i+1=j=1}^{\infty} j\mu(j)z^{j-1}\right) - z\cdot\left(\sum_{i=1}^{\infty} i\mu(i)z^{i-1}\right) \\ &= \psi'_\mu(z)(1-z).\end{aligned}$$

Let us now express the generating function $\psi_\xi(z)$ in two variables $\varphi(\xi,z) = \psi_\xi(z)$. Since $dw_\xi = \frac{d\xi}{1-\xi}$, the above differential equation gives rise to the partial differential equation

$$(1-\xi)D_\xi\varphi(\xi,z) = (1-z)D_z\varphi(\xi,z)$$

with boundary condition $\varphi(0,z) = \psi_\mu(z)$. It is easily verified that this p.d.e. has unique solution

$$\varphi(\xi,z) = \psi_\mu(z+\xi-z\cdot\xi) = \psi_\mu(\xi+(1-\xi)z),$$

which yields an alternate, and perhaps simpler, way to establish the same result.

Part III

Structure

Chapter 6

The 2-Core and the Giant Component

At this point, our technical foundations are fully developed, and we may begin to put our methodology to work. Our first two applications will involve structures related to the connectivity of the random graph $\boldsymbol{\mathcal{G}}(\mu)$:

- the 2-*core*, which is the maximal induced sub-graph with minimum degree at least 2;
- the *giant component*, which is a connected component "much larger" than all the others.

The 2-core and the giant component are natural entry points to the theory of random graphs, and the algorithmic method in particular. This is partly due to the fact that these, while not at all trivial, these are among the easiest random graph problems to solve. Indeed, devising an algorithm to find either the 2-core or the largest connected component is entirely straightforward, and the algorithms involved lend themselves naturally to analysis by differential equations. With the machinery we now have in place, the solutions will come rather smoothly, and this will help ease the transition from abstract methodology to applications.

Another reason for starting of with these particular properties is that they provide useful high-level insight into the structure of the random graph $\boldsymbol{\mathcal{G}}(\mu)$. As we shall see, the "interesting" part of $\boldsymbol{\mathcal{G}}(\mu)$ is the part that belongs to both the 2-core and the giant component. In future applications, a thorough understanding of both of these will often allow us to move directly to the structural heart of $\boldsymbol{\mathcal{G}}(\mu)$, where the more challenging questions lie.

6.0.5 Chapter Organization

The 2-core is analyzed in section 6.1, and the decomposition into connected components in section 6.2. In 6.3, we derive some additional results about the structure of sub-critical random graphs, which are random graphs $\boldsymbol{\mathcal{G}}(\mu)$ for which $M(\mu) < 1$. In this case, all of the connected components have size $o(m)$, and the structure of $\boldsymbol{\mathcal{G}}(\mu)$ is somewhat "trivial."

6.0.6 History and Background

The giant component for $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$ was analyzed in the first papers of Erdős and Rényi [28, 29]. For the configuration model, the decomposition into connected components was first determined by Molloy and Reed [55, 56], using the same basic techniques as appear in this chapter.

The 2-core problem is a special case of the k-core problem, which is analyzed in chapter 9. We have separated the presentation of the 2-core due to its significance with respect to the local structure and the diameter of $\boldsymbol{\mathcal{G}}(\mu)$, which we will analyze in the next to chapters.

The k-core problem for the Erdős-Rényi random graphs $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$ was first solved by Pittel, Spencer, and Wormald [59]. The k-core problem for general degree distributions was solved independently by Molloy [54], Cooper [22], and Janson and Luczak [38], and Fernholz and Ramachandran [32], all using similar techniques.

The presentation of the 2-core result in this chapter differs somewhat from that in [32]; the use of topological representation, differential equations, and the probability generating function simplifies the analysis substantially. Our present analysis of the 2-core, and our reproduction of the giant component results of Molloy and Reed, both emphasize the use of the probability generating function in ways which resemble the use of the p.g.f. in the theory of branching branching processes [4].

6.1 The 2-Core

The 2-core of a graph is the maximal induced sub-graph with minimum degree at least 2; an illustrative example is shown in figure 6.1. In this section, we determine the size and residual distribution of the 2-core of a random graph $\boldsymbol{\mathcal{G}}(\mu)$.

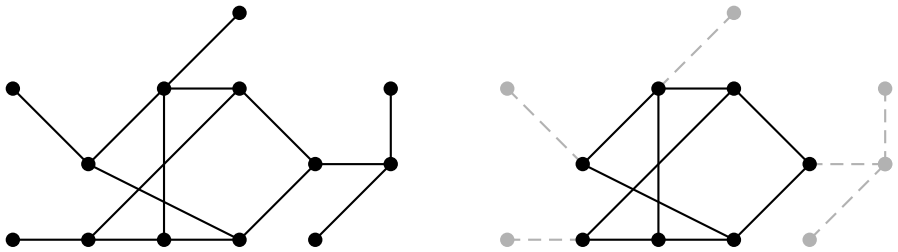


Figure 6.1: A graph and its 2-core.

We shall formally define the 2-core of a configuration $G = (A, V, E)$ in terms of its endpoint set. First we recall some notation and definitions from chapter 3:

- $\vec{E}(a)$ denotes the endpoint matched to a;
- an endpoint set $B \subseteq A$ is *edge-closed* if, for any $a \in B$, $\vec{E}(a) \in B$ as well;
- for $a \in B$, $\mathrm{res}_B(a) = |B(V(a))\backslash\{a\}|$ is the residual degree with respect to (B, V), which counts the number of other endpoints in B which belong to the same vertex as a.

The 2-core of a configuration can now be defined as follows.

Definition 6.1.1. The *2-core* endpoint set of $G = (A, V, E)$ is the maximal edge-closed subset $B \subseteq A$ satisfying $\mathrm{res}_B(a) \geq 1$ for all $a \in B$.

Since the properties of being edge-closed and having minimum residual degree at least 1 are both preserved under the taking of unions, then the 2-core endpoint set exists and is unique for every G (though it may be empty). Moreover, since residual degree is one

less than true degree, the condition that $\mathrm{res}_B(a) \geq 1$ ensures that, for any $v \in V$, either $\deg_B(v) = 0$ or $\deg_B(v) \geq 2$.

If B is the 2-core endpoint set, we may similarly define the 2-core edge set by $E(B)$, and, since B is edge-closed, then this edge set will be a perfect matching of B. The actual 2-core will thus be the configuration $(B, V(B), E(B))$, where $V(B) = \{V(a) : a \in B\}$ is the set of all vertices which contain at least one endpoint in B.

We are interested in both the size and the residual distribution of the 2-core, which we shall describe using the probability generating function (p.g.f.). We briefly review some notation related to the p.g.f.:

- $\psi_\mu(z) = \sum_{i=0}^{\infty} z^i \mu(i)$ is the p.g.f. of a distribution μ;
- z_μ denotes the smallest fixed point in $[0, 1]$;
- if $\mu(1) < 1$ then there at most two such fixed points, one of which is $\psi_\mu(1) = 1$;
- $z_\mu < 1$ if and only if either $M(\mu) > 1$ or $\mu(1) = 1$.

Our main result is as follows.

Theorem 6.1.1. *The following hold w.e.h.p. for a random graph $\boldsymbol{\mathcal{G}}(\mu)$ such that $\mu(1) < 1$:*

1. *the size of the 2-core, as a fraction of the endpoint set, converges to $(1 - z_\mu)^2$;*
2. *if $z_\mu < 1$, the residual distribution of the 2-core converges to the distribution generated by*

$$\varphi(z) = \frac{1 - \psi_\mu(1 - z - z \cdot z_\mu)}{1 - z_\mu}. \tag{6.1}$$

Since $z_\mu < 1$ if and only if $M(\mu) > 1$ (assuming $\mu(1) < 1$), it immediately follows that, the size of the 2-core is $o(m)$ if $M(\mu) \leq 1$ and $\Omega(m)$ if $M(\mu) > 1$.

6.1.1 The 2-core Process

The 2-core of any graph can be found by a simple greedy algorithm: choose an edge incident on a vertex of degree 1, remove this edge, and continue until no such edges remain. We shall analyze this algorithm by analogously defining a constructive configuration model (CM) algorithm.

The basic constructive CM algorithm was described in section 3.2 of chapter 3; briefly, the iterative version involves removing one endpoint from an endpoint partition at each of m time steps in such a way that the matching

$$\mathcal{E}(\mathbf{a}_1, \ldots, \mathbf{a}_m) = \{\{\mathbf{a}_1, \mathbf{a}_2\}, \ldots, \{\mathbf{a}_{m-1}, \mathbf{a}_m\}\}$$

is uniformly random. The analysis of such an algorithm is based an the *endpoint removal process*, which is the descending chain of subsets $(\mathbf{A}_0 \supseteq \cdots \supseteq \mathbf{A}_m)$ given by $\mathbf{A}_{t+1} = \mathbf{A}_t \backslash \{\mathbf{a}_{t+1}\}$.

Algorithm 6.1.1. *A* CM 2-core process *is an endpoint removal process satisfying the following:*

1. *if t is* even*:*
   (a) *if $\mathbf{A}_t$ contains no endpoints of residual degree* 0*, set $\boldsymbol{\tau} = t$ and terminate;*
   (b) *otherwise, choose any $\mathbf{a}_{t+1} \in \mathbf{A}_t$ with residual degree* 0*, and set $\mathbf{A}_{t+1} = \mathbf{A}_t \backslash \{\mathbf{a}_{t+1}\}$;*
2. *if t is* odd*, choose $\mathbf{a}_{t+1}$ uniformly at random from $\mathbf{A}_t$ and set $\mathbf{A}_{t+1} = \mathbf{A}_t \backslash \{\mathbf{a}_{t+1}\}$;*

Note that the way $\mathbf{a}_{t+1}$ is chosen from $\mathbf{A}_t$ for even values of t is not determined precisely, and hence this algorithm can be further specialized if necessary.

Proposition 6.1.2. *For a CM* 2*-core process with initial state $H = (A, V)$:*

1. *if $\boldsymbol{\mathcal{E}}(\mathbf{A}_\tau)$ is a uniformly random matching of the terminal endpoint set $\mathbf{A}_\tau$, then*

$$\mathbf{E} = \mathcal{E}(\mathbf{a}_1, \ldots, \mathbf{a}_{\boldsymbol{\tau}}) \cup \boldsymbol{\mathcal{E}}(\mathbf{A}_\tau)$$

is a uniformly random matching of A, and therefore $\mathbf{G} = (A, V, \mathbf{E}) \stackrel{d}{=} \boldsymbol{\mathcal{G}}(H)$;

2. $\mathbf{A}_\tau$ *is the* 2*-core endpoint set of* $\mathbf{G}$*.*

Proof. The fact that $\mathbf{E}$ is uniformly distributed follows from the fact that $\mathbf{a}_{t+1}$ is chosen uniformly at random from $\mathbf{A}_t$ for all odd values of t, as shown in section 3.2 of chapter 3.

For the second claim, termination occurs at the first even time for which $\mathbf{res}_\tau(a) \geq 1$ for all $a \in \mathbf{A}_\tau$, and therefore $\mathbf{A}_\tau$ is contained in the 2-core endpoint set by definition. Now consider any edge-closed set B which properly contains $\mathbf{A}_\tau$, and let t denote the largest t such that $B \subseteq \mathbf{A}_t$. In this case, we must have $t < \boldsymbol{\tau}$ and, since $\mathbf{a}_{t+1}$ is the lowest numbered endpoint which belongs to $\mathbf{A}_t$, then $\mathbf{a}_{t+1} \in B$.

Since B is edge-closed, t must be even, and in this case, we must have $\mathbf{res}_t(\mathbf{a}_{t+1}) = 0$ otherwise the 2-core process would have terminated at this time. Since $B \subseteq \mathbf{A}_t$ then $\mathbf{a}_{t+1} \in B$ and $\mathrm{res}_B(\mathbf{a}_{t+1}) \leq \mathbf{res}_t(\mathbf{a}_{t+1}) = 0$, so B cannot be the 2-core, and the proof is complete. □

6.1.2 Topological Representation

The 2-core process makes use of only two selection methods, namely *deterministic* and *random*. As described in section 5.2 of the previous chapter, we shall represent these selection methods in terms indicator and measure processes. The selection method at time t describes the method by which the next endpoint $\mathbf{a}_{t+1}$ is chosen from $\mathbf{A}_t$, so in this case, we have

- $\Delta\boldsymbol{\sigma}_{t+1}[D] = \mathbf{I}_t[D] = 1$ if $t < \boldsymbol{\tau}$ is even;
- $\Delta\boldsymbol{\sigma}_{t+1}[D] = \mathbf{I}_t[D] = 1$ if $t < \boldsymbol{\tau}$ is odd.

There are two additional important pieces of information. The first is that the stopping time $\boldsymbol{\tau}$ is the first even time at which there are no endpoints of residual degree 0. Hence, the residual distribution at this stopping time must satisfy $\boldsymbol{\mu}_\tau(0) = 0$. The second is that all deterministic selections have true degree 1 (i.e. residual degree 0), and this fact can be expressed as

$$\Delta\boldsymbol{\sigma}_t[D(1)] = \Delta\boldsymbol{\sigma}_t[D]$$

for all t, since the process $(\boldsymbol{\sigma}_t[D(1)])$ counts deterministic selections of true degree 1. We may use this information to derive the following set of solution constraints.

Proposition 6.1.3. *The following constraints hold u.a. for the* 2*-core process:*

1. $d\sigma_\xi[D] = d\sigma_\xi[R] = d\xi/2$ *if* $\xi < \tau$;
2. $\mu_\tau(0) = 0$;
3. $d\sigma_\xi[D(0)] = d\sigma_\xi[D]$, *so* $d\sigma_\xi[D(i)] = 0$ *for* $i > 1$.

Proof. Immediate. □

6.1.3 Solution to the Residual Distribution

Recall that the incremental weight of a step at time t is defined by $\Delta\mathbf{w}_{t+1} = 1/\mathbf{m}_t = 1/(m-t)$, and that $\mathbf{w}_t[R]$ and $\mathbf{w}_t[D]$ denote the total random and deterministic weight at time t. We shall analyze the 2-core process in terms of these weighted totals, and in fact we shall change variables again by defining

$$x_\xi = 1 - e^{-w_\xi[R]}$$
$$y_\xi = 1 - e^{-w_\xi[D]}.$$

In this case, since random and deterministic selections alternate we trivially have unique solutions

$$x_\xi = y_\xi = 1 - e^{-w_\xi/2} = 1 - \sqrt{1-\xi}$$

for $\xi \leq \tau$. However, we shall solve for the residual distribution in terms of the variables x_ξ and y_ξ, since in future applications, it will not always be the case that random and deterministic steps alternate in this way.

Lemma 6.1.4. *For any regular removal process satisfying* $d\sigma_\xi[D(i)] = 0$ *for all* $i > 0$ *and all* ξ *in some interval* $[0, 1-\epsilon]$, *the unique w.e.h.p. solution to* μ_ξ *in this interval has p.g.f.*

$$\psi_{\mu_\xi}(z) = \psi_\xi(z) = \frac{\psi(x_\xi + (1-x_\xi)z) - y_\xi}{1 - y_\xi},$$

where $\psi = \psi_\mu$ is the p.g.f. of the initial distribution μ, and $x_\xi = 1 - e^{-w_\xi[R]}$ and $y_\xi = 1 - e^{-w_\xi[D]}$.

Proof. Due to the condition $d\sigma_\xi[D(i)] = 0$, the values $\mu_\xi(i)$ for $i \geq 1$ are uniquely determined by corollary 5.3.3 to theorem 5.3.1 from the previous chapter. Hence, for any ξ and any $i \in \mathbb{Z}^*$, let us define

$$\begin{aligned}\nu_\xi(i) &= e^{w_\xi[D]} \sum_{j=i}^{\infty} \mu(j) \binom{j}{i} (1 - e^{-w_\xi[R]})^{j-i} e^{-iw_\xi[R]} \\ &= \frac{1}{1 - y_\xi} \sum_{j=i}^{\infty} \mu(j) \binom{j}{i} x_\xi^{j-i} (1 - x_\xi)^i,\end{aligned}$$

noting that the residual distribution must satisfy $\mu_\xi(i) = \nu_\xi(i)$ for all $i \geq 1$.

Now, the weights $\nu_\xi(i)$ do not sum to 1, and hence ν_ξ is not a proper distribution. Nevertheless, we may still define a corresponding generating function, by letting $\varphi_\xi(z) = \psi_{\nu_\xi}(z) = \sum_i \nu_\xi(i) z^i$. Due to proposition 5.4.3 from §5.4.4, this function can be expressed in terms of the original p.g.f. $\psi = \psi_\mu$ by

$$\varphi_\xi(z) = \frac{\psi(x_\xi + (1 - x_\xi)z)}{1 - y_\xi}.$$

Since the value $\mu_\xi(0)$ corresponds to the constant term in the generating function, and since $\mu_\xi(i) = \nu_\xi(i)$ for all $i \geq 1$, it follows that the p.g.f. of the actual residual distribution μ_ξ differs from $\varphi_\xi(z)$ by a constant. This constant can be computed by noting that the weights of the distribution μ_ξ must indeed sum to 1, and therefore the generating function $\psi_\xi = \psi_{\mu_\xi}$ satisfies $\psi_\xi(1) = 1$.

We thus let

$$\psi_\xi(z) = \varphi_\xi(z) - \frac{y_\xi}{1 - y_\xi} = \frac{\psi(x_\xi + (1 - x_\xi)z) - y_\xi}{1 - y_\xi},$$

and verify that $\psi_\xi(1) = \frac{\psi(1) - y_\xi}{1 - y_\xi} = 1$. □

In particular, note that the solution to such a process at time ξ does not depend on exactly when the random and deterministic removals occur, only the total weight of the

random and deterministic removals. Intuitively, this can be understood by decomposing the above functional transformation into two separate transformations

$$\psi(z) \mapsto \psi(z + x - x \cdot z)$$

and

$$\psi(z) \mapsto \frac{\psi(z) - y}{1 - y}.$$

The first of these corresponds to a random selection, as computed in the previous chapter, while the second corresponds to deterministic removal of an endpoint with residual degree 0. Since these transformations commute, it makes no difference which one is applied first. Accordingly, in any simple removal process which only involves these two selection methods, the solution depends only on the total weight of each method, and not the order in which they are performed.

We may denote the combined functional transformation by $\alpha_{(x,y)}$, so

$$(\alpha_{(x,y)}\psi)(z) = \frac{\psi(z + x - x \cdot z) - y}{1 - y}.$$

For the 2-core process, since random and deterministic endpoints alternate, we have $x_\xi = y_\xi = 1 - \sqrt{1 - \xi}$, and hence the solution is given by

$$\psi_\xi(z) = (\alpha_{(x_\xi, x_\xi)}\psi)(z) = \frac{\psi(x_\xi + (1 - x_\xi)z) - x_\xi}{1 - x_\xi}. \tag{6.2}$$

6.1.4 Termination

Lemma 6.1.4 determines the unique solution to μ_ξ for $\xi < \tau$, and to complete the analysis, we must determine when the algorithm terminates.

Lemma 6.1.5. *Assume $\mu(1) < 1$, and let z_μ denote the smallest fixed point in the p.g.f. $\psi = \psi_\mu$. Then the unique w.e.h.p. solution to the stopping time of the 2-core process is given by*

$$x_\tau = (1 - \sqrt{1 - \tau}) = z_\mu.$$

Proof. For any p.g.f. we have $\psi_\mu(0) = \mu(0)$, and therefore, using (6.2), we have

$$\mu_\xi(0) = \psi_\xi(0) = \frac{\psi(x_\xi) - x_\xi}{1 - x_\xi},$$

and since we have the constraint $\mu_\tau(0) = 0$, it is evident that every possible solution to τ must satisfy $\psi(x_\tau) = x_\tau$. In particular, $x_\tau \geq z_\mu$, since z_μ is the by definition the smallest such fixed point.

Moreover, if $\psi(x_\xi) < x_\xi$ for some $\xi < \tau$ then the residual distribution would have to satisfy $\mu_\xi(0) < 0$, which is impossible, and hence the stopping time must also satisfy

$$x_\tau \leq \inf\{x \in [0,1] : \psi(x) < x\}.$$

Therefore if $\mu(1) < 1$, we have $x_\tau \leq z_\mu$, and the only possible solution is $x_\tau = z_\mu$. □

Theorem 6.1.1 now follows more or less immediately from the lemmas proved above.

Proof of Theorem 6.1.1. The size of the 2-core is given by $m - \boldsymbol{\tau}$, and as shown above, the stopping time has w.e.h.p. solution $x_\tau = z_\mu$, and hence $\tau = 1 - (1 - z_\mu)^2$. Hence, the size of the 2-core converges w.e.h.p. to $(1 - z_\mu)^2$, as a fraction of the endpoint set. Similarly, the residual distribution of the 2-core has unique solution μ_τ, which is thus generated by

$$\psi_\tau(z) = \frac{\psi(z_\mu + (1 - z_\mu)z) - z_\mu}{1 - z_\mu}.$$

□

6.1.5 A Visual Interpretation

The solution to the 2-core process can be understood visually by plotting the p.g.f. $\psi = \psi_\mu$ of the initial distribution μ. The functional transformation

$$\psi(z) \mapsto (\alpha_{(x,y)}\psi)(z) = \frac{\psi(x + z - x \cdot z) - y}{1 - y},$$

which describes the execution of this process corresponds to "moving the origin" from the point $(0, 0)$ to the (x, y), as shown in figure 6.2.

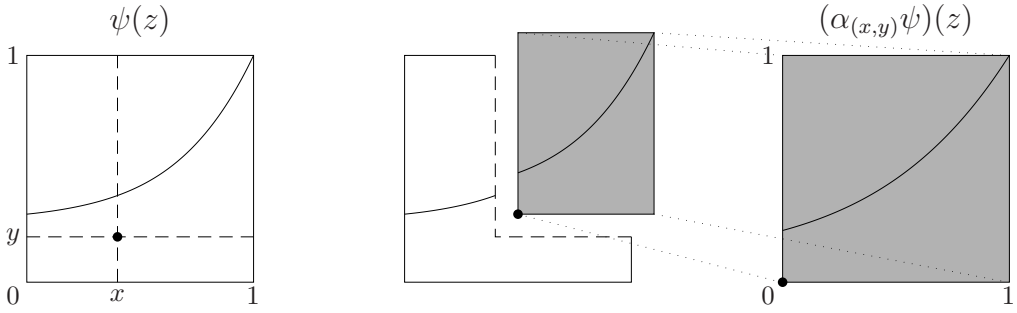


Figure 6.2: The function $(\alpha_{(x,y)}\psi)(z)$ can be plotted be by moving the origin to (x, y) and rescaling the unit square as shown.

During the 2-core process, we have $y_\xi = x_\xi$, so as time passes, the point $(x_\xi, y_\xi) = (x_\xi, x_\xi)$ moves in a diagonal line from $(0, 0)$ to $(1, 1)$, until eventually $\psi(x_\xi) = x_\xi$, at which point the process must terminate since otherwise we would have $\psi_\xi(0) < 0$. This is shown in figure 6.3.

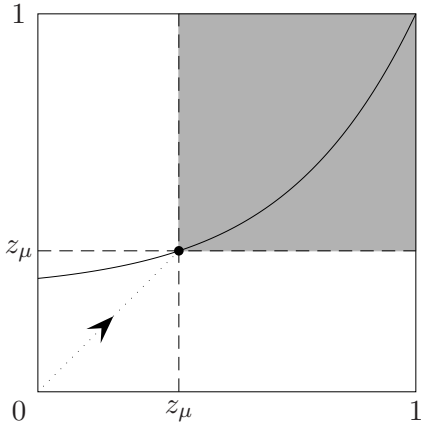


Figure 6.3: The solution to the 2-core process: the point (x_ξ, y_ξ) moves diagonally until the fixed point z_μ is reached. The p.g.f. of the 2-core is plotted in the shaded square, and the fraction of the original endpoint set which belongs to the 2-core is given by the surface area $(1 - z_\mu)^2$ of the shaded square.

6.2 Connected Components

We now discuss the decomposition of $\boldsymbol{\mathcal{G}}(\mu)$ into connected components, and in particular, we address the question of whether $\boldsymbol{\mathcal{G}}(\mu)$ has a *giant component*, which is a component

containing $\Omega(m)$ endpoints.

As with the 2-core, we shall define connected components in terms of endpoint subsets. Recall that an *edge-closed* endpoint set, as defined above, is an endpoint set such that if $a \in B$ then $\vec{E}(a) \in B$ as well. Similarly:

- $\vec{V}(a) = A(V(a))\backslash\{a\}$ is the set of other endpoints which belong to the same vertex as a;
- an endpoint set $B \subseteq A$ is *vertex-closed* if, for any $a \in B$, $\vec{V}(a) \subseteq B$.

Definition 6.2.1. A *connected component* is a minimal non-empty endpoint set which is both edge-closed and vertex-closed.

It is easily verified that this definition is equivalent to the standard notion of a connected component in a graph, which one exception: an isolated vertex (i.e. a vertex which contains no endpoints) does not qualify as a connected component by the above definition. For practical purposes, though, this is merely an accounting detail, which can be easily resolved if the need arises.

Our main result regarding connected components is as follows.

Theorem 6.2.1. *If $\mu(1) < 1$, the following hold w.e.h.p. for the random graph $\boldsymbol{\mathcal{G}}(\mu)$:*

1. *the size of the largest component, as a fraction of the endpoint set, converges to $1 - z_\mu^2$, where z_μ is the smallest fixed point of the p.g.f.;*
2. *the second largest component contains $o(m)$ endpoints.*

In particular, if $z_\mu = 1$, that is, if $M(\mu) \leq 1$ (and $\mu(1) < 1$), then the largest component will also have size $o(m)$. Hence, the existence of a giant component can be determined from the first moment $M(\mu)$ of the residual distribution, and the condition is the same as for the existence of a "giant" 2-core. Our algorithmic analysis will also yield various kinds of additional information about the decomposition into connected components.

Theorem 6.2.2. *If $\mu(1) < 1$, the following hold w.e.h.p. for $\boldsymbol{\mathcal{G}}(\mu)$:*

1. *the number of connected components, as a fraction of the endpoint set, converges to*

$$\int_{x=0}^{z_\mu} (\psi_\mu(x) - x)\, dx;$$

2. *if $z_\mu = 1$, then the average number of vertices per component converges w.e.h.p. to*

$$\frac{2}{2 - M(\lambda)},$$

where λ is the corresponding degree distribution under the assumption that $\lambda(0) = 0$;

3. *if $z_\mu > 0$, then the residual distribution of the sub-graph with the giant component removed converges to the distribution generated by*

$$\varphi(z) = \frac{\psi_\mu(z \cdot z_\mu)}{z_\mu}.$$

6.2.1 The Search Process

Connected components in a graph can be found by executing a standard greedy search, and this algorithm can be expressed as a constructive CM process as follows.

Algorithm 6.2.1. *A* CM search process *is an endpoint removal process satisfying the following.*

- *The state consists of an endpoint partition $\mathbf{H}_t = (\mathbf{A}_t, V)$, along with:*
  - *a subset $\mathbf{Q}_t \subseteq \mathbf{A}_t$ called the* endpoint queue*, which is initially empty;*
  - *an integer $\mathbf{C}_t$ called the* component count*, which is initially $\mathbf{C}_0 = 0$;*
- *At each time step $0 \leq t < m$:*
  1. *if t is* even *and $\mathbf{Q}_t \neq \varnothing$, choose any $\mathbf{a}_{t+1} \in \mathbf{Q}_t$;*
  2. *if t is* even *and $\mathbf{Q}_t = \varnothing$:*
     (a) *increment the component count by $\mathbf{C}_{t+1} = \mathbf{C}_t + 1$;*

(b) *choose* $\mathbf{a}_{t+1}$ *uniformly at random from* $\mathbf{A}_t$*;*

3. *if* t *is* odd*, choose* $\mathbf{a}_{t+1}$ *uniformly at random from* $\mathbf{A}_t$*;*

4. *in all cases, set* $\mathbf{A}_{t+1} = \mathbf{A}_t \backslash \{\mathbf{a}_{t+1}\}$ *and update the queue by:*

   (a) *if* $\mathbf{a}_{t+1} \in \mathbf{Q}_t$*, set* $\mathbf{Q}_{t+1} = \mathbf{Q}_t \backslash \{\mathbf{a}_{t+1}\}$*;*

   (b) *if* $\mathbf{a}_{t+1} \notin \mathbf{Q}_t$*, set* $\mathbf{Q}_{t+1} = \mathbf{Q}_t \cup \vec{V}(\mathbf{a}_t)$*;*

Proposition 6.2.3. *For a CM search process with initial state* $H = (A, V)$*:*

1. *the edge set* $\mathbf{E} = \mathcal{E}(\mathbf{a}_1, \ldots, \mathbf{a}_m)$ *is uniformly random, and therefore* $\mathbf{G} = (A, V, \mathbf{E}) \overset{d}{=} \boldsymbol{\mathcal{G}}(H)$*;*

2. *any two endpoints* $\mathbf{a}_s, \mathbf{a}_t$ *belong to the same connected component of* $\mathbf{G}$ *if and only if* $\mathbf{C}_s = \mathbf{C}_t$*, and therefore:*

   (a) *the number of components is the terminal component count* $\mathbf{C}_m$*;*

   (b) *the size of the largest component (in endpoints) is equal to the longest time interval during which the component count remains constant.*

Proof. The fact that $\mathbf{E}$ is uniformly distributed again follows from the fact that $\mathbf{a}_{t+1}$ is chosen uniformly at random from $\mathbf{A}_t$ for all odd values of t.

For the second claim, let $\boldsymbol{\tau}$ denote the last time that $\mathbf{C}_\tau = 1$, and note that, by induction, it suffices to show that the set $\mathbf{R} = \{\mathbf{a}_1, \ldots, \mathbf{a}_{\boldsymbol{\tau}}\}$ is the endpoint set for a single connected component. It is clear that $\boldsymbol{\tau}$ must be even, and therefore $\mathbf{R}$ is edge-closed. Also, since it must be the case that $\mathbf{Q}_\tau = \varnothing$, then all endpoints added to the queue before time $\boldsymbol{\tau}$ belong to $\mathbf{R}$. The final step (4b) thus ensures that $\mathbf{R}$ is vertex-closed as well.

It only remains to show that the set $\mathbf{R}$ is a single connected component. Hence, choose the smallest t such that $\mathbf{a}_{t+1}$ does not belong to the same component as $\mathbf{a}_1$. Clearly t must be even, otherwise $\mathbf{a}_t$ would belong to the same component a $\mathbf{a}_{t+1}$. Also, $\mathbf{a}_{t+1}$ must be the first endpoint from the vertex $V(\mathbf{a}_{t+1})$ added to the queue. But this can only occur if $\mathbf{Q}_t = \varnothing$, which means that $\mathbf{C}_{t+1} = \mathbf{C}_t + 1$, and therefore $\mathbf{a}_{t+1}$ does not belong to $\mathbf{R}$. □

6.2.2 Topological Representation

The topological representation and solution to the search process are slightly more complicated than for the 2-core process. First, we have to account for the additional state information, in the form of the component count and the endpoint queue. Second, random and deterministic steps no longer alternate, since a random step occurs whenever the endpoint queue becomes empty. As a result, we shall use a slightly different representation for the search process.

The way to deal with the component count is clear: we treat $\mathbf{C}_t$ as if it were a measure process, which means dividing by the total number of steps m, so a solution in this case is a function (C_ξ) such that $\mathbf{C}_{\lfloor \xi m \rfloor} \to C_\xi$ for all $\xi \in [0,1]$. Moreover, increments of the component count satisfy $\Delta \mathbf{C}_{t+1} = 1$ if and only if t is even and $|\mathbf{Q}_t| = 0$, and in this case a random selection, rather than a deterministic selection occurs. Hence, for any even value of t, we have the equation

$$\mathbf{C}_t = \boldsymbol{\sigma}_t[R] - t/2.$$

We will similarly scale the endpoint queue by letting

$$\mathbf{q}_t = |\mathbf{Q}_t| / \mathbf{m}_t$$

denote the fraction of unexposed endpoints which belong to $\mathbf{Q}_t$. However, unlike the component count, $\mathbf{q}_t$ does not satisfy a Lipschitz condition: although $|\mathbf{Q}_t|$ can only decrease by 1 each step, if an endpoint belonging to a vertex of high degree is added to the queue, $|\mathbf{Q}_t|$ will increase by a correspondingly large amount.

As a result, we cannot use Azuma's inequality in its literal form to derive differential constraints on $\mathbf{q}_t$. While it is not difficult to get around this, we will take an alternate route, motivated by the observation that, once an endpoint has been added to $\mathbf{Q}_t$, we are no longer interested in its residual degree. The unexposed residual distribution $\boldsymbol{\mu}_t$ is therefore not appropriate, since it does not distinguish between endpoints based on membership in the endpoint queue.

We shall call an endpoint *unexplored* if it does not belong to the queue, and accordingly define the *unexplored residual distribution* by

$$\boldsymbol{\nu}_t(i) = \frac{|\{a \in \mathbf{A}_t \backslash \mathbf{Q}_t : \mathbf{res}_t(a) = i\}|}{\mathbf{m}_t}.$$

Note that we are dividing by the *total* number of remaining (i.e. unexposed) endpoints $\mathbf{m}_t = |\mathbf{A}_t|$, and not just unexplored endpoints. Hence, $\boldsymbol{\nu}_t$ is not a proper probability distribution, since $\|\boldsymbol{\nu}_t\|_1 = \sum_i \boldsymbol{\nu}_t(i)$ may not be equal to 1.

In fact, the fraction of endpoints in $\mathbf{A}_t$ which are not taken into account by $\boldsymbol{\nu}_t$ is precisely $\mathbf{q}_t = \frac{|\mathbf{Q}_t|}{\mathbf{m}_t}$, and thus we have

$$\mathbf{q}_t = 1 - \|\boldsymbol{\nu}_t\|_1 .$$

This equation does not immediately translate to a solution constraint, since it involves an infinite summation, but it is easy to show that $\boldsymbol{\nu}_t$ is summable. Using the above observations, we may deduce the following solution constraints.

Lemma 6.2.4. *The following solution constraints hold w.e.h.p.*[1] *for all* $\xi \in [0, 1)$*:*

1. $C_\xi = \sigma_\xi[R] - \xi/2$*, and this extends by continuity to* $\xi = 1$*,*
2. $q_\xi = 1 - \|\nu_\xi\|_1$*,*
3. $dC_\xi = 0$ *whenever* $q_\xi > 0$ *(equivalently, whenever* $\|\nu_\xi\|_1 < 1$*),*
4. *for each* $i \in \mathbb{Z}^*$*:*

$$d\nu_\xi(i) = \nu_\xi(i) \left(\frac{d\sigma_\xi[D] - i \cdot d\sigma_\xi[R]}{1 - \xi} \right). \tag{6.3}$$

Proof. The first constraint is immediate since $\mathbf{C}_t = \boldsymbol{\sigma}_t[R] - t/2$ for all even t as noted above. For the next three constraints, since $\boldsymbol{\nu}_t$ is not identical to the residual distribution $\boldsymbol{\mu}_t$, it is necessary to first show that $\boldsymbol{\nu}_\xi = \boldsymbol{\nu}_{\lfloor \xi m \rfloor}$ satisfies the regularity conditions of theorem 5.1.1, but this is straightforward.

[1] The first three of these constraints actually hold u.a., but the differential constraint only holds w.e.h.p., so this stronger guarantee is of no consequence.

First, since at most $(i+1)$ endpoints of residual degree i can become explored in any given step, then $|\Delta(\mathbf{m}_t \cdot \boldsymbol{\nu}_t(i))| < (i+1)$, and this implies that $\boldsymbol{\nu}_t$ is Lipschitz for any fixed interval $t/m \in [0, 1-\epsilon]$. Next, we note that $\boldsymbol{\nu}_\xi(i) \leq \boldsymbol{\mu}_\xi(i)$ for all i and all ξ, since $\boldsymbol{\mu}_\xi(i)$ also counts endpoints which belong to the queue. Since $\boldsymbol{\mu}_\xi$ is uniformly summable on $\xi \in [0, 1-\epsilon]$ by theorem 5.1.1 then $\boldsymbol{\nu}_\xi$ is also uniformly summable by dominated convergence.

Summability of $\boldsymbol{\nu}_\xi$ implies that the equation $\mathbf{q}_t = 1 - \|\boldsymbol{\nu}_t\|_1$ yields the above solution constraint. Moreover, uniform summability implies that $\mathbf{q}_\xi$ is asymptotically equicontinuous; hence, $\mathbf{q}_\xi$ will converge uniformly to any solution, and therefore, if $\mathbf{q}_\xi \to q_\xi > 0$ then we must have $\mathbf{q}_\zeta > 0$ in some neighborhood $\zeta \in (\xi - \epsilon, \xi + \epsilon)$, which means that the component count cannot be incremented in this interval.

For the differential constraint, as noted above, the total number of unexplored endpoints of residual degree i is simply $\mathbf{m}_t \cdot \boldsymbol{\nu}_t(i)$. Whenever an endpoint $\mathbf{a}_{t+1}$ is removed, this value remains unchanged unless $\mathbf{a}_{t+1}$ is unexplored and has residual degree i. In this case, all $(i+1)$ endpoints belonging to the same vertex are added to the queue, and we have $\Delta\left(\mathbf{m}_{t+1} \cdot \boldsymbol{\nu}_{t+1}(i)\right) = -(i+1)$.

Since deterministically chosen endpoints always belong to the queue, this can only occur during random selection, and the probability of choosing such an endpoint uniformly at random is precisely $\boldsymbol{\nu}_t(i)$. We thus have

$$\mathsf{E}_t[\Delta\left(\mathbf{m}_{t+1} \cdot \boldsymbol{\nu}_{t+1}(i)\right)] = -(i+1)\boldsymbol{\nu}_t(i).$$

The quantity $\mathbf{m}_t = m - t$ scales to $(1-\xi)$, and thus by martingale concentration we have the w.e.h.p. differential constraint

$$d((1-\xi)\nu_\xi(i)) = -(i+1)\nu_\xi d\sigma_\xi[R].$$

The left hand side is simply $(1-\xi)d\nu_\xi(i) - \nu_\xi(i)d\xi$, and therefore, using the fact that $d\sigma_\xi[D] + d\sigma_\xi[R] = d\xi$, we have

$$\begin{aligned} d\nu_\xi(i) &= \frac{\nu_\xi(i)\left(d\xi - (i+1)d\sigma_\xi[R]\right)}{1-\xi} \\ &= \nu_\xi(i)\left(\frac{d\sigma_\xi[D] - i \cdot d\sigma_\xi[R]}{1-\xi}\right). \end{aligned}$$

□

6.2.3 Solution to the Residual Distribution

As with the 2-core process, we will again change variables and express the solution to ν_ξ in terms of its generating function. In this case, we shall use the variables

$$x_\xi = e^{-w_\xi[R]}$$
$$y_\xi = e^{-w_\xi[D]},$$

where as usual $dw_\xi[R] = \frac{d\sigma_\xi[R]}{1-\xi}$ and $dw_\xi[D] = \frac{d\sigma_\xi[D]}{1-\xi}$. Note that $x_0 = y_0 = 1$, and that x_ξ and y_ξ are both non-increasing with ξ. Moreover, since $\xi = 1 - e^{-w_\xi}$ then we have $x_\xi \cdot y_\xi = 1 - \xi$.

Since the $\nu_\xi(i)$ do not sum to 1, then the generating function $\psi_\xi = \psi_{\nu_\xi}$ will not satisfy $\psi_\xi(1) = 1$. Instead, we have $\psi_\xi(1) = \sum_{i=0}^{\infty} \nu_\xi(i) = \|\nu_\xi\|_1$, and hence the size of the endpoint queue can be computed by

$$q_\xi = 1 - \psi_\xi(1).$$

In terms if the above change of variables, we may now compute the generating function of ν_ξ.

Lemma 6.2.5. *The solution to ν_ξ is generated by*

$$\psi_\xi(z) = \frac{\psi(x_\xi \cdot z)}{y_\xi},$$

where $x_\xi = e^{-w_\xi[R]}$ and $y_\xi = e^{-w_\xi[D]}$, and $\psi = \psi_\mu$ is the p.g.f. of the initial residual distribution μ.

Proof. First, note that the differential equations in (6.3) can be written

$$d\nu_\xi(i) = \nu_\xi(i)\left(dw_\xi[D] - i \cdot dw_\xi[R]\right).$$

Since each coordinate is independent, we may solve each one individually, and the unique

solution is

$$\begin{aligned}\nu_\xi(i) &= e^{w_\xi[D]-iw_\xi[R]}\mu(i)\\ &= x_\xi^i y_\xi^{-1}\mu(i),\end{aligned}$$

where μ is the initial residual distribution (which is also the initial unexplored residual distribution). The corresponding generating function is therefore given by

$$\psi_\xi(z) = \sum_{i=0}^{\infty} x_\xi^i z^i y_\xi^{-1}\nu_\xi(i) = \frac{\psi(x_\xi \cdot z)}{y_\xi}.$$

□

6.2.4 Random and Deterministic Steps

To complete the analysis we must determine the number of random and deterministic steps which take place, since this determines both the component count and the solutions to y_ξ and x_ξ. We shall use the same change of variables for this purpose, so we first compute

$$\begin{aligned}dx_\xi = d(e^{-w_\xi[R]}) &= -x_\xi dw_\xi[R]\\ &= \frac{-x_\xi d\sigma_\xi[R]}{1-\xi}.\end{aligned} \tag{6.4}$$

The same holds for dy_ξ, and thus we have

$$\frac{dy_\xi}{dx_\xi} = \frac{y_\xi}{x_\xi}\left(\frac{d\sigma_\xi[D]}{d\sigma_\xi[R]}\right). \tag{6.5}$$

Also, since at least half of the steps are random, then $dx_\xi < 0$ for all $\xi < 1$ and also $\lim_{\xi\to 1} w_\xi[R] = \infty$, and it follows that that the mapping $\xi \mapsto x_\xi$ is bijective from $[0,1) \to (0,1]$. We may therefore express y_ξ as a function of x_ξ such that

$$y_\xi = y(x_\xi).$$

for all $\xi \in [0,1)$. In the initial state of the algorithm, we have both $y_0 = x_0 = 1$, and therefore this function must satisfy $y(1) = 1$. Using this information, we may compute a solution as follows.

Lemma 6.2.6. *The solutions to x_ξ and y_ξ satisfy*

$$y_\xi = \max\{x_\xi, \psi(x_\xi)\},$$

where ψ is the p.g.f. of the initial residual distribution μ.

Proof. We first recall that the size of the queue can be computed in terms of the p.g.f. by

$$q_\xi = 1 - \psi_\xi(z) = 1 - \frac{\psi(x_\xi)}{y_\xi}.$$

Since the size of the queue cannot be negative, then we immediately have $y_\xi \geq \psi(x_\xi)$. And, since at least half of the steps are random, then $w_\xi[R] \geq w_\xi[D]$, which implies that $y_\xi \geq x_\xi$ as well.

As discussed above, we may express y_ξ as a function of x_ξ, and the above inequalities yield

$$y(x) \geq \max\{x, \psi(x)\}.$$

To show that the opposite inequality must hold as well, we recall that, whenever $q_\xi > 0$, random and deterministic steps alternate. Hence, if $y_\xi = y(x_\xi) > \psi(x_\xi)$, then $d\sigma_\xi[R] = d\sigma_\xi[D]$, and by (6.5), this implies that

$$y'(x) = \frac{y(x)}{x}$$

whenever $y(x) > \psi(x)$.

Let us now assume that $y(z_0) > \max\{z_0, \psi(z_0)\}$ for some $z_0 \in (0, 1)$, and we claim that the set

$$\{x \in [z_0, 1] : y(x) = \max\{x, \psi(x)\},$$

must be empty; this would yield a contradiction, since the function $y(x)$ must satisfy $y(1) = 1 = \psi(1)$. If the above set is non-empty, we may let z_1 denote its infimum, and in this case

$$y'(x) = y(x)/x = y(z_0)/z_0 > 1$$

for $x \in [z_0, z_1)$. It follows that $y(z_1) > z_1$ as well, and therefore the condition $y(z_1) = \max\{z_1, \psi(z_1)\}$ can only be satisfied by $y(z_1) = \psi(z_1) > z_1$.

Conversely, any p.g.f. has a non-decreasing first derivative on the interval $[0, 1)$, which, combined with the fact that $\psi(1) = 1$, implies that we cannot have both $\psi(x) > x$ and $\psi'(x) > 1$ for the same $x \in [0, 1)$. But, if $y(z_1) = \psi(z_1) > z_1$ and $y(x) > \psi(x)$ for $x \in [z_0, z_1)$, then evidently, both $\psi(x) > x$ and $\psi'(x) > y'(x) > 1$ must occur at some single point x in this interval, by continuity and the mean value theorem.

We thus have a contradiction, and therefore the only solution is $y(x) = \max\{x, \psi(x)\}$.

□

6.2.5 The Number of Connected Components

Now that we have a complete solution to the residual distribution for all $\xi \in [0, 1)$, we may begin to extract information about the decomposition of $\boldsymbol{\mathcal{G}}(\mu)$ into connected components. The total number of connected components is equal to the final value of the component count, which can be expressed in terms of random and deterministic steps by

$$C_\xi = \frac{\sigma_\xi[R] - \sigma_\xi[D]}{2}.$$

This information is not directly available from lemma 6.2.6, since x_ξ and y_ξ are functions of the weighted totals $w_\xi[R]$ and $w_\xi[D]$. In order to determine $\sigma_\xi[R]$ and $\sigma_\xi[D]$, we must change variables back again.

Lemma 6.2.7. *The component count satisfies*

$$dC_\xi = \left(\frac{x_\xi \psi'(x_\xi) - \psi(x_\xi)}{2}\right) dx_\xi, \tag{6.6}$$

whenever $0 < x_\xi < \psi(x_\xi)$ *and* $dC_\xi = 0$ *whenever* $\psi(x_\xi) < x_\xi$.

Proof. As computed in (6.4), we have $dx_\xi = \frac{-x_\xi d\sigma_\xi[R]}{1-\xi}$, and $dy_\xi = \frac{-y_\xi d\sigma_\xi[D]}{1-\xi}$, and since $x_\xi \cdot y_\xi = e^{-w_\xi} = 1 - \xi$, we may compute

$$\begin{aligned} dC_\xi &= \frac{1-\xi}{2}\left(\frac{dy_\xi}{y_\xi} - \frac{dx_\xi}{x_\xi}\right) \\ &= \frac{x_\xi dy_\xi - y_\xi dx_\xi}{2}. \end{aligned}$$

By lemma 6.2.6, we have $y_\xi = \psi(x_\xi)$ whenever $\psi(x_\xi) > x_\xi$, in which case

$$\frac{dC_\xi}{dx_\xi} = \frac{x_\xi \psi'(x_\xi) - \psi(x_\xi)}{2}.$$

And, if $\psi(x_\xi) < x_\xi$, then $y_\xi = x_\xi$, in which case $dC_\xi = 0$. □

We can now immediately compute the number of connected components as follows.

Corollary 6.2.8. *If $\mu(1) < 1$, the number of connected components in $\boldsymbol{\mathcal{G}}(\mu)$, as a fraction of m, and excluding isolated vertices, converges w.e.h.p. to*

$$-\frac{z_\mu^2}{2} + \int_{x=0}^{z_\mu} \psi(x) = \int_{x=0}^{z_\mu} (\psi(x) - x)dx. \tag{6.7}$$

where z_μ is the smallest fixed point in the p.g.f. $\psi = \psi_\mu$.

Proof. As suggested by the above equation, we will change variables and compute the integral in terms of x_ξ. Note that $\psi(x) > x$ if $x < z_\mu$, and $\psi(x) < x$ if $x > z_\mu$, so it suffices to consider values of $x_\xi < z_\mu$. Since $x_0 = 1$, and $x_1 = 0$, then we must reverse the sign, and compute

$$\begin{aligned} C_1 = \int_{x_\xi=1}^{0} \left(\frac{dC_\xi}{dx_\xi}\right) dx_\xi &= \int_{x=0}^{z_\mu} \frac{\psi(x) - x_\xi \psi'(x)}{2} dx \\ &= -\frac{z_\mu \psi(z_\mu)}{2} + \int_{x=0}^{z_\mu} \psi(x)dx, \end{aligned}$$

using integration by parts in the last step. Since $z_\mu = \psi(z_\mu)$, then $z_\mu \psi(z_\mu) = z_\mu^2$, and the second expression in (6.7) follows by noting that $\int_{x=0}^{z_\mu} dx = z_\mu/2$ □

Corollary 6.2.9. *If $M(\mu) \leq 1$ and $\mu(1) < 1$, and if the degree distribution satisfies $\lambda(0) = 0$ (i.e. if there are no isolated vertices), then the average number of vertices per component converges w.e.h.p. to*

$$\frac{2}{M(\lambda) - 2}.$$

Proof. In §5.4.3 of the previous chapter, the average vertex degree $M(\lambda)$ was expressed in terms of the p.g.f. $\psi = \psi_\mu$ of the residual distribution by

$$M(\lambda) = \left(\int_{z=0}^{1} \psi(z) \right)^{-1},$$

under the assumption that $\lambda(0) = 0$ (i.e. ignoring isolated vertices). If $M(\mu) \leq 1$ and $\mu(1) < 1$, then this distribution is sub-critical, and therefore $z_\mu = 1$, so the number of connected components, as a fraction of the endpoint set, is given by $\frac{1}{M(\lambda)} - \frac{1}{2}$.

Hence, the average number of endpoints per component in this case is

$$\frac{1}{\frac{1}{M(\lambda)} - \frac{1}{2}} = \frac{2M(\lambda)}{2 - M(\lambda)}.$$

To compute the average number of vertices per component, we divide by the ratio m/n of endpoints to vertices, which is again the average degree $M(\lambda)$. □

6.2.6 The Giant Component

In addition to the total number of components, we may also determine whether there exists a giant component, which is a component of size $\Omega(m)$. In this case, we are interested in the longest time interval during the component count is not incremented.

Lemma 6.2.10. *If $\mu(1) < 1$, the size of the largest connected component of $\boldsymbol{\mathcal{G}}(\mu)$, as a fraction of the number of endpoints, converges w.e.h.p. to $1 - z_\mu^2$. Also:*

1. *the second-largest component has size $o(m)$ w.e.h.p.;*
2. *if $\mu(0) > 0$, the residual distribution of the sub-graph of $\boldsymbol{\mathcal{G}}(\mu)$ with the largest component removed converges w.e.h.p. to the distribution generated by*

$$\varphi(z) = \psi(z \cdot z_\mu)/z_\mu.$$

Proof. Recall that the scaled size of the queue $\mathbf{q}_t$ is asymptotically equicontinuous, which implies that if $\mathbf{q}_{\lfloor \xi m \rfloor}$ converges uniformly to the unique solution over any interval $[0, 1 - \epsilon]$.

In particular, if there exists an interval $\xi \in [\zeta_1, \zeta_2]$ such that then the largest component w.e.h.p. has size at least $|\zeta_2 - \zeta_1|$.

Conversely, if $C_{\zeta_1} > C_{\zeta_2}$, then w.e.h.p. the component count must be incremented at least once in this interval, and therefore if $dC_\xi > 0$ for $\xi \in [\zeta_1, \zeta_2]$, then the largest component revealed during this interval w.e.h.p. has size at most ϵ for any fixed $\epsilon > 0$ (and hence, by choosing $\epsilon > 0$ arbitrarily small, the size is $o(m)$).

Now, if $x_\xi < \psi(x_\xi)$ then $y_\xi > x_\xi$, and hence $dC_\xi > 0$ by lemma 6.2.7. Since $x_\xi < \psi(x_\xi)$ whenever $x_\xi < z_\mu$, it follows that all components revealed in this interval have size at $o(m)$ w.e.h.p.

Similarly, $q_\xi > 0$ holds if and only if $x_\xi > \psi(x_\xi)$, which holds whenever $z_\mu < x_\xi < 1$. And, since $y_\xi = x_\xi = 1 - \xi$ in this interval, it follows that $q_\xi > 0$ for all values of $0 < \xi < 1 - z_\mu^2$, and hence for any $\epsilon > 0$, the largest component has size at least $1 - z_\mu^2 - \epsilon$, and by choosing $\epsilon > 0$ arbitrarily small, we conclude that the size of the largest component converges w.e.h.p. to $1 - z_\mu^2$.

Finally, if $\mu(0) > 0$, then $z_\mu > 0$ as well, and hence the number of endpoints which do not belong to the giant component is $\Omega(m)$. Moreover, the giant component becomes completely revealed at the time ξ for which $x_\xi = z_\mu$, and the residual distribution at this time is generated by $\psi_\xi(z) = \psi(z \cdot z_\mu)/z_\mu$. Now, it may be the case that the giant component is not the first component revealed, but even then, the fact that $q_\xi > 0$ when $z_\mu < x_\xi < 1$ implies that all but $o(m)$ of the endpoints exposed during this time interval belong to the giant component, and these do not asymptotically affect the residual distribution of the sub-graph with the giant component removed. □

6.2.7 A Visual Interpretation

As with the 2-core process, the solution trajectory of the search process can also by understood visually by plotting the p.g.f. In this case the relevant transformation is given by

$$\psi(z) \mapsto (\beta_{(x,y)}\psi)(z) = \frac{\psi(x \cdot z)}{y}.$$

This transformation also corresponds to a rescaling of the unit square, except that it is the top right corner $(1, 1)$ which is relocated to the point (x, y), as shown in in figure 6.4.

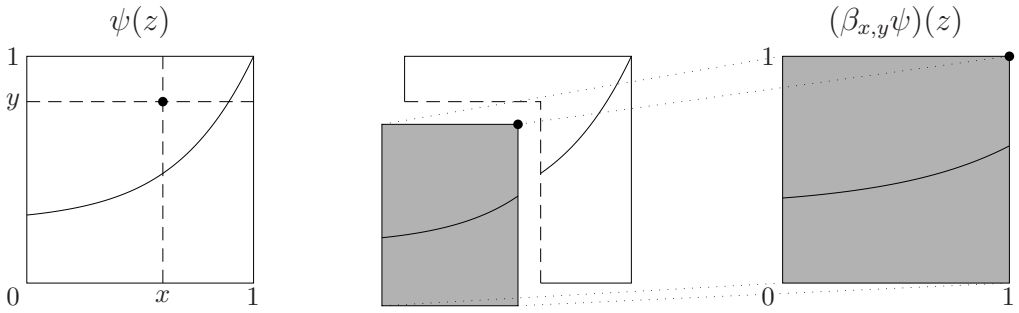


Figure 6.4: The function $(\beta_{x,y}\psi)(z)$ can be plotted by moving the top corner $(1, 1)$ of the unit square to (x, y) and rescaling as shown.

Unlike the 2-core process, the point (x_ξ, y_ξ) does not move in a straight line, but instead follows the path $y_\xi = \max\{x_\xi, \psi(x_\xi)\}$, as shown in figure 6.5.

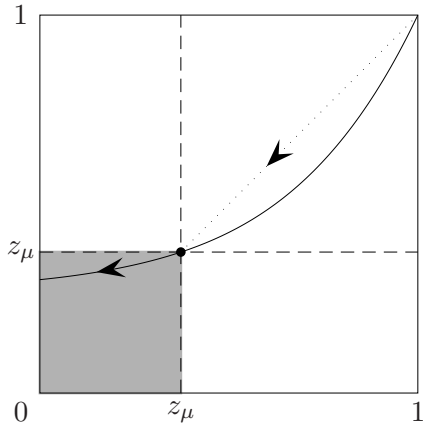


Figure 6.5: The solution to the search process: the point $(x_\xi, y_\xi = x_\xi)$ moves diagonally until the fixed point (z_μ, z_μ) is reached, and then follows the path $y_\xi = \psi(x_\xi)$. The p.g.f. of the graph with the giant component removed is in the shaded square, and the number of endpoints in the giant component is the surface area $1 - z_\mu^2$ of the *unshaded* region.

6.3 The Structure of Sub-Critical Graphs

We now examine the structure of sub-critical random graphs, that is random graphs for which $M(\mu) < 1$, and hence both the 2-core and the giant component contain $o(m)$

endpoints. As we shall see, the structure of such graphs is more or less "trivial."

In order to give a more rigorous description, however, we must use different techniques from those used above. This is because, since each component has size $o(m)$, these components are "invisible" with respect to our typical asymptotic scaling, which involves dividing by the number of total endpoints $|A| = m$. Moreover, the structure of small components can generally not be determined with exponentially high probability, since the number of steps of the CM algorithm necessary to expose such a component is $o(m)$.

Our analysis of sub-critical graphs in this section serves as a useful supplement to the differential equations method. In particular, the differential equations method is generally not useful for analyzing terminal behavior of a process. However, in many cases, the algorithms we are interested will reach a state where the unexposed residual distribution is sub-critical at some time $t = (1 - \epsilon)m$. At this point, the graph which remains will have the "trivial" structure which we describe below. Hence, in general, if we can solve the execution of such an algorithm until the residual distribution becomes sub-critical, we then argue that the terminal behavior of the algorithm will be determined by the general properties of sub-critical graphs.

We will only discuss two properties of sub-critical graphs. The first is the size of the largest connected component, and the second is the number of cycles in any given component. Specifically, we will prove that any connected component in a sub-critical graph will contain at most one cycle; hence every connected component is either a tree or uni-cyclic.

In order to prove these results, we will somewhat arbitrarily impose a bound of $m^{1/8-\epsilon}$ on the maximum degree of the graph. Intuitively, this bound is not overly restrictive, given the fact that the average residual degree is in fact less than 1. Of course, it is possible to construct asymptotic degree sequences for which the average residual degree is less than 1 and yet the maximum degree exceeds $m^{1/8}$. However, in most situations, particularly in the case the configuration model is used to model the Erdős-Rényi random graphs $\mathbf{G}_{n,p}$ and $\mathbf{G}_{n,m}$ with the Poisson distribution, this bound is easily satisfied, and in fact the maximum

degree for Erdős-Rényi graphs is $o(\ln n)$.

We begin with the bound on the maximum component size.

Proposition 6.3.1. *Let $\mathbf{G} = \boldsymbol{\mathcal{G}}(\mu)$ with $M(\mu) < 1$, assume that $M(\mu_{\mathbf{G}}) \to M(\mu)$, and also that the maximum degree of any vertex is at most $m^{1/8-\epsilon}$. Then w.p.h.p. the largest component contains at most $m^{1/4-\epsilon}$ endpoints*

Proof. It suffices to prove that the probability that the first component revealed contains $m^{1/4-\epsilon}$ endpoints is $n^{-1-\Omega(1)}$. Hence, the first $T = m^{1/4-\epsilon} = o(m)$, and since residual distribution changes negligibly during $o(m)$ steps, our analysis will focus on the exact size of the endpoint queue, which we abbreviate by $\mathbf{x}_t = |\mathbf{Q}_t|$.

Now, first note that, during any deterministic step, we have $\Delta\mathbf{x}_{t+1} = -1$, since in this case $\mathbf{a}_{t+1}$ is removed from $\mathbf{Q}_t$. And, during all random steps, $\Delta\mathbf{x}_{t+1} = -1$ if $\mathbf{a}_{t+1} \in \mathbf{Q}_t$, or else

$$\Delta\mathbf{x}_{t+1} = \left|\vec{V}(\mathbf{a}_{t+1})\right| = \mathrm{res}(\mathbf{a}_{t+1}).$$

Hence, the expected change in $|\mathbf{x}|_t$ is bounded above by the average unexposed residual degree. And, since $M(\mu) < 1$, then we may assume that, during the first $o(m)$ steps, the average unexposed residual degree is uniformly bounded above by $1 - \delta$ for small but fixed $\delta > 0$.

It follows that

$$\mathsf{E}_t[\Delta\mathbf{x}_t \mid \mathbf{I}_t[R] = 1] \leq 1 - \delta$$

and $\mathsf{E}_t[\Delta\mathbf{x}_t \mid \mathbf{I}[D] = 1] = -1$, and hence the predicable component of the Doob decomposition satisfies

$$\mathcal{E}\mathbf{x}_t \leq (\boldsymbol{\sigma}_t[R] - \boldsymbol{\sigma}_t[D]) - \delta\boldsymbol{\sigma}_t[R].$$

And, since, for even values of t, we have both $\boldsymbol{\sigma}_t[C] = \frac{\boldsymbol{\sigma}_t[R]-\boldsymbol{\sigma}_t[D]}{2}$ and $\boldsymbol{\sigma}_t[R] \geq t/2$, we may thus bound the number of components from below by

$$2\boldsymbol{\sigma}_t[C] \geq \mathcal{E}\mathbf{x}_t + \delta t/2$$

for $t = o(m)$.

Moreover, since the size of the endpoint queue can change by at most the maximum degree $m^{1/8-\epsilon}$, then the increment size of $\mathbf{x}_t$ are bounded accordingly. Hence, by Azuma's inequality, the martingale component of $\mathbf{x}_t$ satisfies

$$\mathsf{P}[\mathcal{M}\mathbf{x}_t > m^{1/8-\epsilon}C\sqrt{t}] < e^{-C^2/2}.$$

For $T = m^{1/4-\epsilon}$, we may thus compute

$$\mathsf{P}[\mathcal{M}\mathbf{x}_T > m^{1/4-3\epsilon/2}C] < e^{-C^2/2}.$$

and thus

$$\mathsf{P}[\mathcal{M}\mathbf{x}_T > \epsilon_0 T] < e^{-\Omega(m^\epsilon)}$$

for any $\epsilon_0 > 0$.

Since $\mathbf{x}_T = \mathcal{M}\mathbf{x}_T + \mathcal{E}\mathbf{x}_T$, and $\mathbf{x}_T \geq 0$, then $\mathcal{E}\mathbf{x}_T > -\mathcal{M}\mathbf{x}_T$, and hence

$$\mathsf{P}[\mathcal{E}\mathbf{x}_T < -\epsilon_0 T] < e^{-\Omega(m^\epsilon)},$$

and therefore, by choosing $\epsilon_0 < \delta/2$, the probability that the first component is not completely revealed at this time satisfies

$$\mathsf{P}[\boldsymbol{\sigma}_T[C] = 1] < e^{-\Omega(m^\epsilon)}.$$

□

This bound can in turn be used to show that all components are at most uni-cyclic.

Proposition 6.3.2. *Let $\mathbf{G} = \boldsymbol{\mathcal{G}}(\mu)$ with $M(\mu) < 1$, assume that $M(\mu_{\mathbf{G}}) \to M(\mu)$, and also that the maximum degree of any vertex is at most $m^{1/8-\epsilon}$. Then w.p.h.p. every connected component of $\mathbf{G}$ contains at most one cycle.*

Proof. Again it suffices to prove that the probability that the first component contains more than one cycle is $n^{-1-\Omega(1)}$. Now, any connected component contains at most one cycle if

and only if the number of edges is at most equal to the number of vertices, or equivalently, if the number of endpoints is at most twice the number of vertices.

In the search process, the number of vertices in the first component increases whenever $\mathbf{a}_{t+1} \notin \mathbf{Q}_t$. Since deterministic endpoints are chosen from $\mathbf{Q}_t$, this can only occur for random selections. Now, the first selection is random, as is every even numbered selection until the first component is revealed. Hence, we must show that $\mathbf{a}_{t+1} \in \mathbf{Q}_t$ occurs for at most one random selection before the first component is revealed.

We again abbreviate by $\mathbf{x}_t = |\mathbf{Q}_t|$, and let us denote by $\mathbf{y}_t$ the total number of times $\mathbf{a}_{t+1} \in \mathbf{Q}_t$ occurs for a random selection, so $\mathbf{y}_t = 0$ and $\Delta\mathbf{y}_{t+1} = 1$ if and only if $\mathbf{a}_{t+1} \in \mathbf{Q}_t$ and t is odd (so $(t+1)$ is even).

Now, the probability that $\mathbf{a}_{t+1} \in \mathbf{Q}_t$ for a random selection is $|\mathbf{Q}_t|/(m-t)$. Moreover, by the previous proposition, the first component is revealed after at most $m^{1/4-\epsilon}$ time steps. And, since only one endpoint can be removed each step, then this means the queue can only grow to size $m^{1/4-\epsilon}$ during this time w.e.h.p.

To make use of this, let us denote by $\boldsymbol{\tau}$ the minimum of $m^{1/4-\epsilon}$ and the first time the queue size exceeds $m^{1/4-\epsilon}$. Note, then, that the number of cycles in the first component is at most $\mathbf{y}_{\boldsymbol{\tau}}$ w.e.h.p. And, for any $t \leq \boldsymbol{\tau}$, we have

$$\mathsf{P}_t[\Delta\mathbf{y}_{t+1} = 1] < m^{1/4-\epsilon}/(m - m^{1/4-\epsilon}) < m^{-3/4}.$$

We may thus compute

$$\mathsf{P}[\mathbf{y}_{\boldsymbol{\tau}} > 1] \leq \binom{T}{2}(m^{-3/4})^2 \leq T^2/m = m^{-1-2\epsilon}.$$

□

Chapter 7

Local Structure and the Branching Process

In this chapter and the next, we put the topological abstraction developed in chapter 4 on hold, along with the differential equations method, as we study the combinatorial structure of $\boldsymbol{\mathcal{G}}(\mu)$ in the local vicinity of a single vertex.

Our method is still algorithmic: we shall examine the execution of a standard breadth-first search (BFS) on a random graph. However, the BFS algorithm will only execute for $O(\ln n)$ steps, and the global parameters, such as the residual distribution $\boldsymbol{\mu}_t$, will remain essentially unchanged for this short duration. As such, the nature of the analysis will remain combinatorial, and a different set of techniques is required. Moreover, due to this short duration, it is generally not possible to achieve exponentially high guarantees in probability, and instead we shall settle for results which only hold with polynomially high probability (w.p.h.p.).

The objective of the current chapter is to introduce this kind of local analysis, and to draw a somewhat informal connection between BFS on a random graph and a classical kind of random process called a *branching* or *Galton-Watson process*. This informal discussion sets the stage for the analysis of the diameter of $\boldsymbol{\mathcal{G}}(\mu)$ in the next chapter, during which the informal connection between random graphs and branching processes is made more rigorous.

Chapter Organization

We begin in section 8.1 by introducing notation and basic definitions related to breadth-first search on the random graph $\boldsymbol{\mathcal{G}}(\mu)$. We then discuss the qualitative behavior of this process, and we draw an informal connection to a branching process generated by the same residual distribution.

In section 7.2, we give a brief introduction to the theory of branching processes, with emphasis on the use of the probability generating function. In particular, we show how the decomposition of a random graph $\boldsymbol{\mathcal{G}}(\mu)$ based on the 2-core and the giant component corresponds naturally to a similar decomposition of a branching process.

History and Background

The correspondence between the local structure of a random graph and a branching process is well-known; the analysis of the giant component if $\mathbf{G}_{n,p}$ in [3], for example, uses the branching process analogy.

The introduction to branching processes in section 7.2 is both brief and informal. For a much more detailed introduction we refer the reader to a number of texts on the subject, for example [4].

7.1 Neighborhoods and Breadth-First Search

We shall describe the local structure of a configuration $G = (A, V, E)$ in the vicinity of a given vertex $v \in V$ in terms of endpoint subsets. For any vertex v, we recursively define endpoint sets $N_i(v)$ and $R_i(v)$ by:

- $R_0(v) = A(v)$ and $N_0(v) = \varnothing$;
- $N_{i+1}(v) = N_i(v) \cup R_i(v) \cup \vec{E}(R_i(v))$;
- $R_{i+1}(v) = A(V(N_{i+1}(v)))\backslash N_{i+1}(v)$.

The sets $N_i(v)$ and $R_i(v)$ are respectively the i'th *(endpoint) neighborhood* and *perimeter* of v. The i'th endpoint neighborhood contains all endpoints which would be found during a standard *breadth-first search* (BFS), as illustrated in figure 7.1.

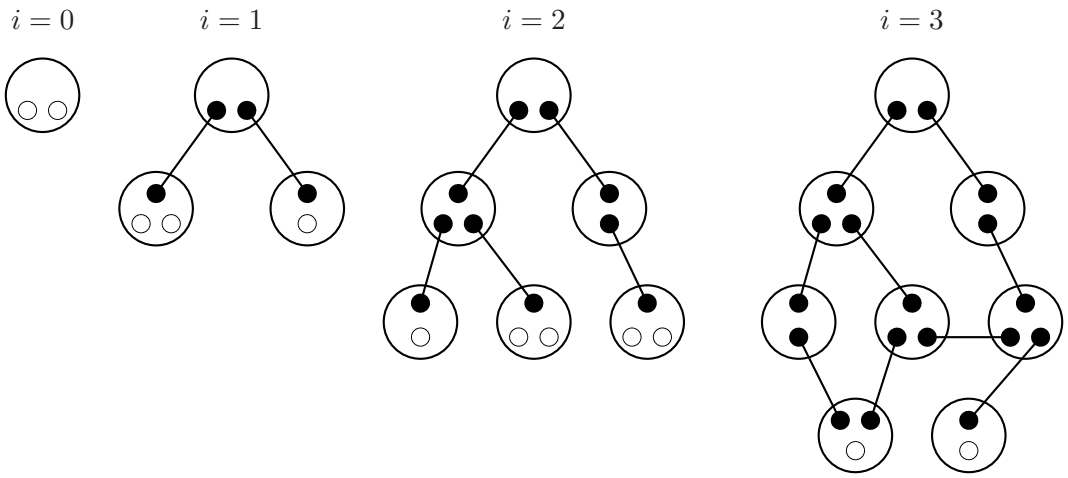


Figure 7.1: Neighborhoods and perimeters of a vertex v for $i = 0, 1, 2, 3$. In each case, neighborhood $N_i(v)$ consists of all matched endpoints (filled in in black), while the perimeter $R_i(v)$ is the set of unmatched endpoints.

7.1.1 The BFS Process

In a random configuration $\boldsymbol{\mathcal{G}}(A, V)$, neighborhoods and perimeters are random subsets of A, which we accordingly denote in boldface. The neighborhoods of a given vertex v can be exposed sequentially by performing a breadth-first search, and this algorithm can be analyzed in concert with the configuration model. BFS will be discussed in detail in the next chapter, so for now we only give some basic definitions and notation.

- the *BFS exposure process* is the ascending chain of edge sets and $\mathbf{E}_t(v) = \mathbf{E}(\mathbf{N}_t(v))$;
- $\mathbf{r}_t(v) = |\mathbf{R}_t(v)|$, and the sequence $(\mathbf{r}_0(v), \mathbf{r}_1(v), \ldots)$ is simply called the *BFS process*.

The BFS process is therefore simply the sequence of perimeter sizes, but for the purpose of computing conditional probabilities, the "state" of the BFS process is considered to be the entire edge set $\mathbf{E}_t(v)$. In technical terms, then, the BFS process $(\mathbf{r}_t(v))$ is considered adapted to the filtration induced by the BFS exposure process $(\mathbf{E}_t(v))$, and thus the conditional probability of a given event L at time t is

$$\mathsf{P}_t[L] = \mathsf{P}_t[L \mid \mathbf{E}_t(v)].$$

It is immediate by definition that the "increment" of the BFS exposure process ($\mathbf{E}_t(v)$) is precisely the set of edges incident on the perimeter $\mathbf{R}_{t-1}(v)$ so we have:

$$\mathbf{F}_t(v) = \mathbf{E}_t(v) \backslash \mathbf{E}_{t-1}(v) = \mathbf{E}(\mathbf{R}_{t-1}), \tag{7.1}$$

In figure 7.1, the edge sets $\mathbf{E}_t(v)$ for $t = i = 0, 1, 2, 3$ are precisely the edges which appear in each of the diagrams shown, and the edge sets $\mathbf{F}_t(v)$ can be determined accordingly.

At any time during BFS, we distinguish three kinds of endpoints

- an endpoint $a \in A$ is *exposed* if it belongs to an exposed edge in $\mathbf{E}_t(v)$;
- the set of *unexposed* endpoints is denoted by $\mathbf{A}_t$;
- an endpoint belonging to $\mathbf{A}_t(v) \backslash \mathbf{R}_t(v)$ is called *unexplored*.

These definitions are the same as for the search process analyzed in section 6.2. Intuitively, unexplored endpoints are endpoints that we are "unaware of," as opposed to endpoints in $\mathbf{R}_t(v)$, which remain unexposed, but have already been found by the local search.

The sequence $(\mathbf{A}_0(v), \mathbf{A}_1(v), \ldots)$ is an endpoint removal process, and it is easy to verify that the conditional uniformity of the unexposed edge set is preserved.

Proposition 7.1.1. *In any BFS process, the edge set* $\mathbf{E} \backslash \mathbf{E}_t(v) = \mathbf{E}(\mathbf{A}_t(v))$ *is a uniformly random matching of* $\mathbf{A}_t(v)$*, conditional on* $\mathbf{E}_t(v)$*.*

Proof. Since the original edge set $\mathbf{E}$ is uniformly random, then every perfect matching E which contains $\mathbf{E}_t(v)$ will occur with equal probability, the set of all such matchings corresponds bijectively to the set of perfect matchings of $\mathbf{A}_t(v)$. □

Despite the fact that conditional uniformity is preserved, and the similarity to the search process from the previous chapter, the analysis of BFS exposure process is of a different nature qualitatively. This is because, as time passes, $\mathbf{R}_t(v)$ typically grows large, and therefore the number of endpoints exposed in a given iteration is not of order $O(1)$. Also, BFS typically lasts $O(\ln m)$ rather than $\Theta(m)$ steps, and its analysis thus requires a different set of tools.

7.1.2 BFS and Branching Processes

We will examine the BFS process rigorously in the next chapter, and for now our intent is simply to introduce the local structure of the random graph $\boldsymbol{\mathcal{G}}(\mu)$ at a more informal level. Let us simplify our notation by omitting the initial vertex and abbreviating by $\mathbf{R}_t(v) = \mathbf{R}_t$.

Now, if the perimeter $\mathbf{R}_t$ is "small" relative to the unexposed endpoint set $\mathbf{A}_t$, then in the "typical" situation, the following will occur:

1. each $a \in \mathbf{R}_t$ will match to an endpoint outside of $\mathbf{R}_t$ (i.e. an unexplored endpoint);
2. no two endpoints in $\mathbf{R}_t$ will match to the same vertex.

In this case, if an endpoint $a \in \mathbf{R}_t$ matches to an endpoint $\vec{\mathbf{E}}(a)$ on a vertex of degree d, then the remaining $d-1$ endpoints are contributed to $\mathbf{R}_{t+1}$. Intuitively, then, each endpoint in $\mathbf{R}_t$ produces $\deg(\vec{\mathbf{E}}(a)) - 1 = \mathrm{res}(\vec{\mathbf{E}}(a))$ children in $\mathbf{R}_{t+1}$.

Moreover, if $\mathbf{R}_t$ is "small," then the residual degrees of the endpoints in $\vec{\mathbf{E}}(a)$ will be "almost" independent random variables, and the distribution of each of these random variables will be approximately given by the residual distribution μ. Thus, modulo various caveats and technicalities, the relationship between the perimeter sizes $\mathbf{r}_t$ and $\mathbf{r}_{t+1}$ can be roughly approximated by

$$\mathbf{r}_t \simeq \sum_{j=1}^{\mathbf{r}_t} \mathbf{x}_{\mu,j}, \tag{7.2}$$

where the $\mathbf{x}_{\mu,j}$ are i.i.d. μ-distributed random variables.

If strict equality were to hold, then (7.2) would define a classical random process called a *branching* or *Galton-Watson (GW)* process. The corresponding random tree structure is a *Galton-Watson tree*, and thus the local structure of a random graph $\boldsymbol{\mathcal{G}}(\mu)$ can informally be described as that of a GW tree.

Of course, a random graph is not actually a GW tree, and the major technical impediments to the direct application of branching process theory to random graphs are the following:

- we only have asymptotic knowledge of the actual residual distribution of the random graph $\boldsymbol{\mathcal{G}}(H) = \boldsymbol{\mathcal{G}}(\mu)$, in the form of the convergence $\mu_H \to \mu$;
- the residual degrees of successive randomly selected endpoints are not truly independent;
- the residual degree of $\vec{\mathbf{E}}(a)$ for any $a \in \mathbf{R}_t$ does not always accurately count the number of endpoints contributed to $\mathbf{R}_{t+1}$, due the possibility of *cross-edges* of the following kinds:
  - a *horizontal edge* is an edge connecting two endpoints in $\mathbf{R}_t$;
  - a set of *diagonal edges* is a set of edges which connect two or more endpoints in $\mathbf{R}_t$ to the same vertex.

Examples of both horizontal and diagonal edges are shown in figure 7.1: in the diagram for $i = 3$, the two edges incident on the bottom left vertex are diagonal, while the edge connecting the second and third vertices in the second row from the bottom (i.e. the "horizontal" edge) is horizontal.

In our analysis of the diameter of $\boldsymbol{\mathcal{G}}(\mu)$ in the next chapter, a substantial amount of technical effort is dedicated to overcoming these technicalities. Ultimately, none of these will prove fatal, and the diameter of $\boldsymbol{\mathcal{G}}(\mu)$ turns out to be what one would expect based on the branching process analogy.

In a broader sense, though, the extent to which the local similarities between random graphs and GW trees can be made rigorous remains an open question. If we restrict our attention to a finite neighborhood of a vertex v, then this connection can be made rigorous fairly easily. For instance, given any fixed constant T, the fact that the distribution of the finite sequence $(\mathbf{r}_0(v), \ldots, \mathbf{r}_T(v))$ for an arbitrary vertex v converges to that of a branching process based on the residual distribution μ is more or less immediate by definition. However, the information we can glean from finite neighborhoods of this sort is limited in value, and the real question is whether some generally correspondence between random graph properties and GW tree properties can be established.

7.2 A Brief Introduction to Branching Processes

For a distribution μ on $\mathbb{Z}^*$, a *branching process* generated by μ is a random process $(\mathbf{y}_t)$ which satisfies

$$\mathbf{y}_{t+1} \stackrel{d}{=} \sum_{i=0}^{\mathbf{y}_t} \mathbf{x}_{\mu,i}, \tag{7.3}$$

where the $\mathbf{x}_{\mu,i}$ are i.i.d. μ-distributed. Branching process are classical and well-understood probabilistic structures with various applications (see, e.g. [4]), the simplest of which is as a model for reproduction and population growth. We imagine an initial population of $\mathbf{y}_0$ organisms, each of which produces an independent and μ-distributed number of offspring. In this case, the total number of offspring will be distributed according to $\mathbf{y}_1 = \sum_{i=0}^{\mathbf{y}_t} \mathbf{x}_{\mu,i}$, and similarly the size of the t'th generation is given by $\mathbf{y}_t$.

A branching process is simply a random sequence of non-negative integers, but the population metaphor of independent reproduction also gives rise to a corresponding random tree called a *Galton-Watson tree*. A GW-tree can be defined recursively as follows. We begin with a single root vertex $\mathbf{v}_0$, and we assign this vertex a random number of children according to the distribution μ. Then, we recursively generate a GW tree rooted at each of these children. A Galton-Watson tree thus represents the genealogy tree of descendants of a given organism, in which the parent-child relationship between vertices literally describes the biological relationship between the corresponding organisms.

In this section, we give a brief introduction to branching processes, and we explore the informal connection between the local structure of a random graph $\boldsymbol{\mathcal{G}}(\mu)$ and the corresponding GW tree. We shall omit proofs and focus on intuition; for a more rigorous introduction, we refer the reader to [4].

7.2.1 The Probability Generating Function

The probability generating function (p.g.f.) was introduced in the previous chapter, and served a critical role in our methods for studying the 2-core and search algorithms on a random graph. The p.g.f. is also an essential tool in the theory of branching processes,

and the fact that the p.g.f. is instrumental in analyzing both kinds of structures is certainly not "coincidence," given the close relationship between the two. However, the standard methods for using the p.g.f. to study branching processes differ, at least superficially, from the methods developed in the previous chapters for analyzing algorithms on random graphs.

The basic observation relating the p.g.f. to a branching process is the following. For any distribution μ, let us denote by μ_t the distribution of the t'th generation $\mathbf{y}_t$ for a μ-generated branching process with initial population size $\mathbf{y}_0 = 1$ (and note that $\mu_1 = \mu$). Now, it is straightforward, though somewhat laborious, to write down the explicit equation for μ_t. On the other hand, the p.g.f. $\psi_t = \psi_{\mu_t}$ can be described quite easily.

The p.g.f. can be expressed in "probabilistic" notation as $\psi_\mu(z) = \mathsf{E}[z^{\mathbf{x}_\mu}]$, where $\mathbf{x}_\mu$ is μ-distributed, and the p.g.f. for the sum of k independent, μ-distributed random variables can similarly be computed by

$$\psi(z) = \mathsf{E}[z^{\sum_{i=1}^{k} \mathbf{x}_{\mu,i}}] = \prod_{1=0}^{k} \mathsf{E}[z^{\mathbf{x}_{\mu,i}}] = \psi(z)^k.$$

Given a ν-distributed random variable $\mathbf{y}$, the sum of $\mathbf{y}$ such i.i.d. random variables thus has p.g.f.

$$\sum_{i \in \mathbb{Z}^*} \nu(i) \psi_\mu(z)^i = \psi_\nu \circ \psi_\mu(z).$$

It follows that the generating functions ψ_t for the distributions $\mu_t = \mathfrak{D}[\mathbf{y}_t]$ in a μ-generated branching process are given by $\psi_t(z) = \psi_{t-1} \circ \psi_\mu(z)$, and since $\psi_1 = \psi_\mu$, then ψ_t is in fact the t'th iterate of original p.g.f. ψ_μ:

$$\psi_t(z) = \psi_\mu \underbrace{\circ \cdots \circ}_{t} \psi_\mu(z). \tag{7.4}$$

There are many implications to the fact that the generations of a branching process correspond to iterations of the generating function, and we will not discuss these in detail. At a qualitative level, though, we may observe that stable points under iteration correspond to fixed points of ψ_μ. And, as discussed in the previous chapter, unless $\mu(1) = 1$, then ψ_μ has at most two fixed points in the interval $[0, 1]$: the point 1, and, if $M(\mu) > 1$, the second

fixed point $z_\mu < 1$. Moreover, since ψ_μ has non-negative derivatives of all orders, then $\psi(z) > z$ if $0 \leq z < z_\mu$, and $\psi(z) < z$ if $z_\mu < z < 1$.

It follows that, for any $z \in [0, 1)$, we have $\lim_{t\to\infty} \psi_t(z) = z_\mu$; as such, the fixed point z_μ plays an important role in the analysis of branching processes, which, as we shall see, mirrors the role of the fixed point in the in the analysis of random graphs.

7.2.2 Survival and Extinction

One of the simplest and most important events related to a branching process is *extinction*, which is the event

$$\bigvee_{t\in\mathbb{N}} [\mathbf{y}_t = 0].$$

The complement of extinction is survival, which occurs if $\mathbf{y}_t > 0$ for all t. As the name suggests, extinction occurs if the number of descendants of the the initial organisms is finite.

One method for computing the extinction probability is to iterate the p.g.f. as described above. For any distribution μ, we have $\mathsf{P}[\mathbf{x}_\mu = 0] = \mu(0) = \psi_\mu(0)$, so in a branching process, we have

$$\lim_{t\to\infty} \mathsf{P}[\mathbf{y}_t = 0] = \lim_{t\to\infty} \psi_t(0) = z_\mu,$$

and hence the extinction probability is equal to z_μ.

The extinction probability can also be computed recursively, without explicitly iterating the p.g.f., by noting that the extinction event for the root vertex of a GW tree is simply the conjunction of the extinction events for each of its children. Since these sub-trees are independent, it follows that the extinction probability satisfies

$$p = \mathsf{E}[\textstyle\prod_{i=1}^{\mathbf{x}_\mu} p] = \mathsf{E}[p^{\mathbf{x}_\mu}] = \psi_\mu(p).$$

Hence, the extinction probability must be a fixed point in the p.g.f., and while this argument does not immediately determine which of the two possible fixed points is the actual extinction probability, it is not hard to show that the correct fixed point is the smallest fixed point z_μ. More significantly, this method of computing the extinction probability ex-

emplifies a typical technique for analyzing branching processes and GW trees by exploiting their recursive symmetry.

7.2.3 Sizes of the Giant Component and the 2-core

As shown in the previous chapter, the existence of a fixed point z_μ which is strictly smaller than 1 signals the existence of a giant component (and a giant 2-core) in a random graph $\boldsymbol{\mathcal{G}}(\mu)$. Hence, in this sense an infinite GW tree corresponds to a both a giant component and a giant 2-core in a random graph.

However, the extinction probability for a GW is exactly equal to the fixed point z_μ, and while the existence of such a fixed point corresponds to the existence of both a giant component and a giant 2-core, neither of these have size $1 - z_\mu$ (as a fraction of the total endpoint set). This accounting detail, though, is easily resolved.

The fraction of all endpoints belonging to the 2-core or the giant component is evidently the same as the fraction of all edges belong to either structure. Since survival of a branching process corresponds to a BFS search tree which grows "large," then each edge in fact has two chances to survive and reach the giant component.

Accordingly, the probability of extinction in both directions is z_μ^2, which is in line with the size of the giant component $1 - z_\mu^2$ computed in the previous chapter. In order for an edge to belong to the 2-core, survival in both directions is required, which yields the correct formula $(1 - z_\mu)^2$ for the fraction of all edges in the 2-core.

7.2.4 Conditional Distributions

The relationship between survival of a branching process and both the giant component and the 2-core extends beyond just the extinction probability. To illustrate this, we now consider the structure of a GW tree conditional on either survival or extinction.

Extinction and Small Components

We first consider a GW tree conditional on extinction, and our first observation that this structure will also be GW generated by a different distribution μ_{Ext}; this is because the extinction of the root tree is equivalent to the extinction of all of the sub-trees, and hence, conditional on extinction of the entire tree, each of these sub-trees will be i.i.d., extinction-conditioned GW trees.

In order to compute the distribution μ_{Ext}, note that the extinction probability, conditional on the event that the root vertex has degree i, is simply z_μ^i. By Bayes Theorem, the probability that the root has degree i, conditional on extinction, is thus given

$$\mu_{Ext}(i) = \frac{\mu(i) z_\mu^i}{z_\mu}.$$

The generating function for this distribution is given by

$$\psi_{Ext}(z) = \frac{\sum_{i=0}^{\infty} \mu(i) z^i z_\mu^i}{z_\mu} = \frac{\psi_\mu(z \cdot z_\mu)}{z_\mu}. \tag{7.5}$$

We have already encountered the above p.g.f. in the previous chapter in our analysis of the giant component of $\boldsymbol{\mathcal{G}}(\mu)$: by Theorem 6.2.2, this function generates the residual distribution of the sub-graph of $\boldsymbol{\mathcal{G}}(\mu)$ with the giant component removed. In other words, as one would expect, the structure of the small components in $\boldsymbol{\mathcal{G}}(\mu)$ resembles the structure of a GW tree conditioned on extinction.

The Skeleton and Survival

The same exact argument does not work for conditioning on survival, because a survival-conditioned GW tree will not in fact be a GW tree. This is because, in order for the root tree to survive, only one of its children must produce an infinite lineage.

Hence, we shall instead consider what is called the *skeleton*, which is the sub-tree consisting only of vertices with surviving lineages. It is not difficult to see that the skeleton must in fact be GW tree, since each surviving child will be independently distributed according to the survival-conditioned distribution.

Let us therefore denote by μ_{Skel} the distribution corresponding to the number of surviving children in a survival-conditioned GW tree. The simplest way to compute μ_{Skel} is in to compute its p.g.f. directly. Hence, let us first denote by $\mathbf{w}$ the number of children in an ordinary GW tree with produce infinite lineages, and attempt to compute the p.g.f. $\mathsf{E}[z^{\mathbf{w}}]$.

Since each child survives independently with probability $(1 - z_\mu)$., then $\mathbf{w}$ can be expressed as $\sum_{i=1}^{\mathbf{x}_\mu} \mathbf{I}_i$, where the $\mathbf{I}_i$ are independent Bernoulli random variables with $\mathsf{P}[\mathbf{I}_i = 1] = (1 - z_\mu)$. Such a Bernoulli random variable has p.g.f. $\mathsf{E}[\mathbf{I}_i] = z_\mu + (1 - z_\mu) \cdot z$, and therefore this generating function is given by

$$\varphi(z) = \mathsf{E}[z^{\mathbf{w}}] = \psi_\mu(z_\mu + (1 - z_\mu) \cdot z).$$

Since extinction corresponds to the event $\mathbf{w} = 0$, or equivalently $z^{\mathbf{w}} = 1$, then the contribution to $\mathsf{E}[z^{\mathbf{w}}]$ by all cases where $\mathbf{w} > 0$ is given by

$$\sum_{i=1}^{\infty} \mathsf{P}[\mathbf{w} = i] \cdot z^i = \varphi(z) - \varphi(0) = \psi_\mu(z_\mu + (1 - z_\mu) \cdot z) - \psi_\mu(z_\mu).$$

Hence, the p.g.f. for the distribution μ_{Skel} of $\mathbf{w}$ conditional on survival is given by

$$\psi_{Skel}(z) = \frac{\psi_\mu(z_\mu + (1 - z_\mu) \cdot z) - \psi_\mu(z_\mu)}{\mathsf{P}[\mathbf{w} > 0]} = \frac{\psi_\mu(z_\mu + (1 - z_\mu) \cdot z) - z_\mu}{1 - z_\mu}. \tag{7.6}$$

As shown in the previous chapter, this equation describes the p.g.f. of the 2-core of $\boldsymbol{\mathcal{G}}(\mu)$. Hence, just as the small components of a random graph have the local structure of an extinction-conditioned GW tree, the 2-core has the structure of the skeleton of a survival-conditioned GW tree.

7.2.5 Structural Relationships and Generating Functions

When considered jointly, the generating functions for μ_{Ext} and μ_{Skel} yield additional information about the relationship between the corresponding sub-structures of either a GW tree or the random graph $\boldsymbol{\mathcal{G}}(\mu)$. Recall that both of these generating functions can be understood visually by plotting the p.g.f., as shown in figure 7.2.

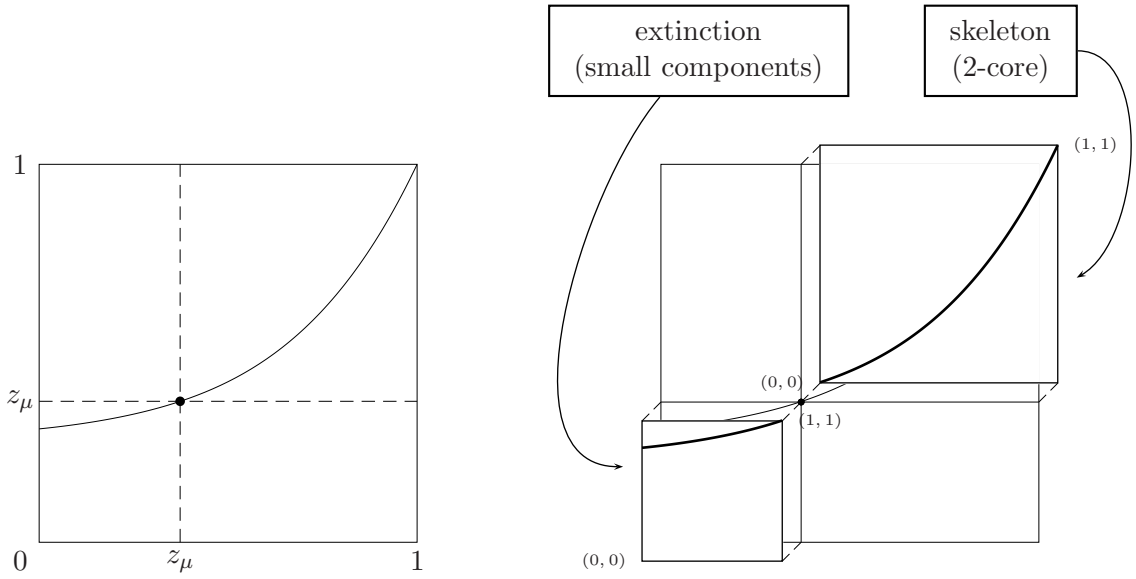


Figure 7.2: The decomposition of a GW tree based on survival and extinction. From the original p.g.f. ψ_μ, shown on the left, we may compute the p.g.f. for both the extinction-conditioned GW tree and the skeleton of a survival-conditioned tree, as shown on the right. For a random graph, these distributions correspond to the small components and the 2-core, respectively.

Recall also that the derivatives of a p.g.f. at 0 and 1 respectively yield the probability weights and factorial moments of a distribution, by

$$\mu(i) = \frac{\psi^{(i)}(0)}{i!},$$
$$M_k(\mu) = \psi^{(k)}(1).$$

Since the fixed point (z_μ, z_μ) represents the origin $(0, 0)$ for the skeleton p.g.f., and the point $(1, 1)$ for the extinction p.g.f, the relationship between these three distributions can be described as follows.

Proposition 7.2.1. *For any distribution μ with $0 < z_\mu < 1$*

1. $\mu_{Ext}(i) = z_\mu^{i-1}\mu(i)$;
2. $z_\mu^{1-k} M_k(\mu_{Ext}) = (1 - z_\mu)^{1-k}\mu_{Skel}(k)$;
3. $M_k(\mu_{Skel}) = (1 - z_\mu)^{1-k} M_k(\mu_{Skel})$.

In particular, the first moment satisfies

$$M(\mu_{Ext}) = \mu_{Skel}(1),$$
$$M(\mu_{Skel}) = M(\mu).$$

Proof. Given that

$$\psi_{Ext}(z) = \frac{\psi_\mu(z \cdot z_\mu)}{z_\mu},$$
$$\psi_{Skel}(z) = \frac{\psi(z_\mu + (1 - z_\mu) \cdot z) - z_\mu}{1 - z_\mu},$$

we may easily compute

$$\psi^{(i)}_{Ext}(z) = z_\mu^{i-1}\psi^{(i)}(z \cdot z_\mu),$$
$$\psi^{(i)}_{Surv}(z) = (1 - z_\mu)^{i-1}\psi^{(i)}(z_\mu + (1 - z_\mu) \cdot z_\mu),$$

and the above are immediate. □

Chapter 8

The Diameter

The *diameter* of a graph is the maximum distance between two vertices in the same connected component. In this chapter, we derive an expression of the form

$$c \ln n \pm o(\ln n)$$

for the diameter of the random graph $\boldsymbol{\mathcal{G}}(\mu)$, where the constant c depends on the limiting residual distribution μ.

The proof of the diameter result involves tracing the rate of neighborhood expansion during the breadth-first search (BFS) process which was introduced in the previous chapter. This sort of local analysis differs qualitatively from global analysis performed in chapter 6, and as a result we will not make use of the differential equations method, and we will often work with explicit constants rather than using the topological approach described in chapter 4.

As a result, the diameter result only holds with polynomially, rather than exponentially high probability. This reflects the nature of the problem rather than the methodology: the diameter is somewhat sensitive to small structural changes in a random graph, and therefore it is not possible to achieve w.e.h.p. concentration.

Since neighborhood sizes during BFS typically grow exponentially, it is not difficult to see that the diameter of $\boldsymbol{\mathcal{G}}(\mu)$ will generally have order of magnitude $\Theta(\ln n)$. However, in order to precisely compute the leading constant, an more detailed analysis of BFS is required. We shall achieve this higher level of precision by exploiting a simple structural relationship between shortest paths and the 2-core of a graph. As such, our argument thus relies on the results about the size and degree distribution of the 2-core from the chapter 6.

Chapter Organization

In section 8.1 we review the basic notation for the BFS process, and we give a summary of our approach to computing the diameter. In section 8.2, we state our main theorem, along with an intuitive explanation of the formula for the diameter of $\mathcal{G}(\mu)$, and an overview the challenges involved in making this intuition rigorous.

The proof of the diameter theorem appears in the subsequent six sections. We prove the upper bound and lower bound separately. The upper bound is the more technical of the two, and is spread over sections 8.3 through 8.6. The lower bound proof appears in section 8.7, and in section 8.8 we discuss certain generalizations and exceptional cases. Finally in section 8.9 we compute the diameter for some specific degree distributions, including the Poisson distribution which corresponds to the Erdős-Rényi random graph $\mathbf{G}_{n,p}$, and the "power-law" graph model introduced by Aiello, Chung, and Lu [2].

History and Background

As a relatively natural and fundamental graph property, the diameter of random graphs has received a fair amount of attention from researchers over time. A number of diameter results are available for denser ranges of the Erdős-Rényi random graph $\mathbf{G}_{n,p}$; these are summarized in [13].

The diameter results of this chapter, which originally appeared in [33], were the first results of asymptotic precision for most sparse random graphs, including sparse $\mathbf{G}_{n,p}$, and the sparse "power-law" model introduced in [2].

Diameter results of such precision for sparse random graphs were known earlier only for regular graphs of constant degree [15]. Weaker results (to within a constant factor) were known for the diameter of sparse $\mathbf{G}_{n,p}$ [20] and random "expected-degree" power-law graphs [50].

8.1 Computing the Diameter of $\boldsymbol{\mathcal{G}}(\mu)$

In this section, we give an overview of our approach to computing the diameter of a random graph $\boldsymbol{\mathcal{G}}(\mu)$. We begin with some basic definitions.

For our purposes, a *path* in a configuration $G = (A, V, E)$ is an even sequence of endpoints $(a_1, \ldots, a_{2j})$ such that:

- $\{a_i, a_{i+1}\} \in E$ for all odd values of i;
- $V(a_i) = V(a_{i+1})$ for all even values of i;
- the *length* of this path is one half of the number of endpoints.

Although our formal definition involves endpoints, we may consider either a vertex or an edge to lie along a path in the obvious sense, and we note that the length of a path is equal to both the number of edges and vertices (counting multiplicities).

We say a path *connects* two vertices u, v if the first endpoint belongs to u and the last endpoint belongs to v. The *distance* between a pair of vertices u, v, which we denote by $\delta(u, v)$ is the length of a shortest path connecting u and v. If u and v belong to different connected components, then we let $\delta(u, v) = \infty$, since in this case there are no paths connecting u and v. Accordingly, the *diameter* of G is the maximum finite distance between any pair of vertices:

$$\Delta(G) = \max\{\delta(u, v) : u \neq v \in V, \delta(u, v) < \infty\}.$$

In other words, the diameter is the length of a longest shortest path in G.

8.1.1 Distances and Breadth-First Search

We shall analyze the diameter by examining the breadth-first search (BFS) process introduced in the previous chapter, and we briefly review our notation. For any vertex v in a configuration $G = (A, V, E)$, we recursively define endpoint sets $N_i(v)$ and $R_i(v)$ by:

- $R_0(v) = A(v)$ and $N_0(v) = \varnothing$;
- $N_{i+1}(v) = N_i(v) \cup R_i(v) \cup \vec{E}(R_i(v))$;
- $R_{i+1}(v) = A(V(N_{i+1}(v)))\backslash N_{i+1}(v)$.

The sets $N_i(v)$ and $R_i(v)$ are respectively the i'th *(endpoint) neighborhood* and *perimeter* of v.

Note that $N_i(v)$ is then the set of all endpoint which belong to a path which begins in $A(v)$, and has length at most i (i.e. contains at most $2i$ endpoints). Moreover, $R_i(v) \cap N_i(v) = \varnothing$, and (for $i > 0$) $R_i(v)$ contains all endpoints belong to vertices in $V(N_i(v))$, but which are not themselves (as endpoints) reachable by such a path.

The distance between two vertices can also be expressed in terms of these endpoint sets by

$$\delta(u, v) = \inf\{i : A(u) \cap N_i(v) \neq \varnothing\}.$$

Accordingly, the distance between two vertices can be understood as the number of BFS iterations which need to be performed, starting from v, until an endpoint belonging to u is encountered (or vice-versa).

In a random configuration $\boldsymbol{\mathcal{G}}(A, V) = (A, V, \mathbf{E})$, these endpoint subsets are random, and we the *BFS exposure process* is the random ascending chain of edge sets

$$(\mathbf{E}_0(v), \mathbf{E}_1(v), \ldots)$$

where $\mathbf{E}_t(v) = 0$ and $\mathbf{E}_{t+1} = \mathbf{E}(\mathbf{N}_t(v))$. Note also, that the "increments" of the BFS exposure process are thus given by

$$\mathbf{E}_{t+1}(v)\backslash\mathbf{E}_t(v) = \mathbf{E}(\mathbf{R}_t(v)).$$

The actual *BFS process* is the sequence $(\mathbf{r}_0(v), \mathbf{r}_1(v), \ldots)$ of perimeter sizes $\mathbf{r}_t(v) = |\mathbf{R}_t(v)|$. For the purposes of conditional probabilities, the "state" at time t is considered to be the entire edge set $\mathbf{E}_t(v)$. At a technical level, then, the BFS process $(\mathbf{r}_t(v))$ is considered adapted to the filtration induced by the BFS exposure process, so, for example,

$$\mathsf{E}_t[\mathbf{r}_{t+1}(v)] = \mathsf{E}[\mathbf{r}_{t+1}(v) \mid \mathbf{E}_t(v)].$$

8.1.2 BFS Hitting Times

Given vertices u, v in a random graph G, perhaps the simplest method for computing the distance $\delta(u, v)$ is to perform to perform BFS beginning at v until the vertex u is encountered. One disadvantage of this naive method, though, is that our search will generally explore a large portion (i.e. a constant fraction of the vertex or endpoint sets) of the graph before finding the second vertex u.

On a random graph, a simple modification of this naive strategy allows us to determine the distance between u and v while only exposing $o(m)$ endpoints. The distance from u to v in any graph G can also be expressed as

$$\delta(u, v) = \inf\{i + j + 1 : \vec{E}(R_{v,i}) \cap R_{u,j} \neq \varnothing\},$$

and thus a shortest path between u and v can be found by performing BFS simultaneously from u and v until we find two endpoint perimeters which are connected by an edge.

In a random graph, our state of knowledge after respectively performing i and j iterations of BFS from u and v consists of the edge sets $\mathbf{E}_i(v)$ and $\mathbf{E}_j(v)$, and at this point the remaining unexposed endpoints will be matched uniformly at random. As in the so-called "birthday paradox," the expected number of edges connecting two endpoint sets of size $m^{1/2}$ in a uniformly random matching is $\Theta(1)$. And, if we choose a slightly larger endpoint size, then we can be virtually assured that such an edge exists.

For any vertex v, let us define the random variable

$$\boldsymbol{\tau}(v, r) = \inf\{t : \mathbf{r}_t(v) \geq r\}, \tag{8.1}$$

noting that this is the hitting time of the the event $\mathbf{r}_t(v) \geq r$ and is therefore an optional time with respect to the BFS edge exposure process. In order to implement the above strategy for finding shortest paths, we shall, somewhat arbitrarily, choose a value "slightly" larger than $m^{1/2}$ by letting

$$\chi = m^{1/2} \ln^2 m,$$

and we now have the following method for bounding distances.

Proposition 8.1.1. *For any random graph $\boldsymbol{\mathcal{G}}$, the inequality*

$$\boldsymbol{\delta}(u, v) \leq \boldsymbol{\tau}(u, \chi) + \boldsymbol{\tau}(v, \chi) + 1 \tag{8.2}$$

holds w.p.h.p. simultaneously for all pairs $u, v \in V$.

Before we offer a proof, we make some brief remarks. First, it may be the case that $\mathbf{r}_t(v)$ never reaches size χ, which means that $\boldsymbol{\tau}(v, \chi) = \infty$. If this occurs, the above proposition is evidently not useful in bounding distances from v to any other vertex. Second, while (8.2) yields an upper bound to $\boldsymbol{\delta}(u, v)$, the opposite inequality does not hold, and therefore it is not clear how this proposition helps establish a lower bound on the diameter.

As we shall see, though, $\boldsymbol{\tau}(u, \chi) < \infty$ for every vertex in the giant component, and for the majority of vertices the sum $\boldsymbol{\tau}(u, \chi) + \boldsymbol{\tau}(v, \chi)$ will not exceed the actual distance $\boldsymbol{\delta}(u, v)$ by more than $o(\ln m)$. In essence, then, the diameter of $\boldsymbol{\mathcal{G}}(\mu)$ is fairly well approximated by

$$2 \cdot \max\{\boldsymbol{\tau}(v, \chi) : v \in V, \boldsymbol{\tau}(v, \chi) < \infty\}. \tag{8.3}$$

Proof of Proposition 8.1.1

The proof of proposition 8.1.1 follows from the same argument as the "birthday paradox."

Lemma 8.1.2. *Consider disjoint subsets $S_1, S_2 \subseteq A$ such that both $|S_1|, |S_2| > m^{1/2} \ln^2 m$. Then*

$$\mathsf{P}[\vec{\mathbf{E}}(S_1) \cap S_2 = \emptyset] = m^{-\omega(1)}.$$

Proof. First, note that any given endpoint in S_1 matches into S_2 with probability at least $|S_2|/m \geq m^{-1/2}$. Now, if a particular endpoint in S_1 does not match into S_2, it may match to another endpoint in S_1. Nevertheless, if we sequentially expose the matches all of the endpoints in S_1, there are at least $|S_1|/2$ chances to find a connection to S_2. The probability that no connection is found is therefore at most

$$(1-m^{-1/2})^{\frac{m^{1/2}\ln^2 m}{2}} = \left(\left(1 - \frac{1}{m^{1/2}}\right)^{m^{1/2}}\right)^{\frac{\ln^2 m}{2}} = \left(\frac{1}{e - o(1)}\right)^{\Omega(\ln^2 m)} = m^{-\Omega(\ln m)} = m^{-\omega(1)}.$$

□

Proof of Proposition 8.1.1. It is an immediate consequence of the above lemma that $\mathsf{P}[\boldsymbol{\delta}(u,v) < \boldsymbol{\tau}(v,\chi)+\boldsymbol{\tau}(v,\chi)+1] = 1-m^{-\omega(1)}$, for every pair of vertices, and since there are $O(m^2)$ such pairs, then these events occur simultaneously with probability $1-m^{\omega(1)}$, which is (more than) polynomially high. □

8.1.3 Shortest Paths and the 2-core

The 2-*core* of a graph, which was examined in detail in chapter 6, is the maximal sub-graph with minimum degree at least 2. The 2-core of a configuration $G = (A, V, E)$ was formally defined in terms of its endpoint set, which we shall denote by A_{2C}, and which is the maximal endpoint set which satisfies the following:

1. A_{2C} is edge closed, meaning that if $a \in A_{2C}$ then $\vec{E}(a) \in A_{2C}$ also,
2. $\mathrm{res}_{2C}(a) > 0$ for all $a \in A_{2C}$, meaning that, for any $a \in A_{2C}$, there exists a distinct $a' \neq a$ in A_{2C} such that $V(a) = V(a')$.

We shall make extensive use of the following variant of the 2-core

Definition 8.1.1. For any $G = (A, V, E)$ and any $B \subseteq A$, the B-*augmented* 2-*core* endpoint set $A_{2C,B}$ is the maximal endpoint set which is edge closed, and such that every endpoint $a \in A_{2C,B} \backslash B$ has induced residual degree $\mathrm{res}_{A_{2C,B}}(a)$ at least 1.

Intuitively, then all endpoints belonging to B are "exempt" from the minimum degree requirement of the 2-core. The significance of the augmented 2-core with respect to path lengths is largely due to the following proposition.

Proposition 8.1.3. *Every path which begins at b_1 and terminates at b_2 is contained in the 2-core endpoint set $A_{2C,\{b_1,b_2\}}$.*

Proof. Let $(a_1, \ldots, a_{2j})$ denote any such path, so $b_1 = a_1$ and $b_2 = a_{2j}$, and let A_0 denote the set $\{a_i\}_{i=1}^{2j}$. By definition, A_0 is edge-closed, and also, every endpoint in A_0, other

than $a_1 = b_1$ and $a_{2j} = b_2$ belongs to the same vertex as either its successor or predecessor along this path. Hence, every endpoint in A_0, other than b_1 and b_2, has induced residual degree $\mathrm{res}_{A_0}(a) \geq 1$, and since the augmented 2-core is the maximal such endpoint set, then $A_0 \subseteq A_{2C,\{b_1,b_2\}}$. □

For an endpoint set B of size $o(m)$, it is not difficult to show that the augmented 2-core will satisfy the same asymptotic properties as the actual 2-core. A proposition to this effect is stated below, but first we review the essential results about the 2-core from chapter 6.

In theorem 6.1.1, it was shown that, if $M(\mu) > 1$, then the 2-core w.e.h.p. has size $\Omega(m)$, and that the residual distribution of the 2-core converges w.e.h.p. to a distribution which can be computed from μ_{2C}. We shall denote the 2-core residual distribution by μ_{2C}, and in Theorem 6.1.1, this distribution was specified in terms of its generating function

$$\psi_{2C}(z) = \frac{1 - \psi(1 - z - z \cdot z_\mu)}{1 - z_\mu},$$

where z_μ is the smallest fixed point of the original p.g.f. $\psi = \psi_\mu$ in the interval $[0, 1]$; the assumption $M(\mu) > 1$ ensures that $z_\mu < 1$, so this is well-defined.

As discussed in the previous chapter, the 2-core distribution μ_{2C} is the same as the skeleton distribution μ_{Skel} for the corresponding branching process. In proposition 7.2.1 using the formula for this generating function, it was shown that $\mu_{2C} = \mu_{Skel}$ satisfies the following properties:

$$\mu_{2C}(0) = \psi'(z_\mu)$$
$$M(\mu_{2C}) = M(\mu) = \psi'(1).$$

Structure of the Augmented 2-core

The augmented 2-core behaves quite similarly to the actual 2-core, and so most of the analysis from chapter 6 is applicable to the augmented 2-core without much modification. Recall that the 2-core of $\boldsymbol{\mathcal{G}}(A, V)$ can be found by successively removing an endpoint of

residual degree 0 followed by a uniformly random endpoint from A until no endpoints of residual degree 0 remain. What remains at this time is the 2-core endpoint set, and since these endpoints are unexposed, then the 2-core edge set is uniformly random, conditional in the 2-core endpoint set.

The B-augmented 2-core can be found in the same way, except that now the termination condition is that every endpoint outside of B must have residual degree at least 1; as noted above, endpoints in B are exempt from this restriction. The conditional uniformity property thus evidently holds in this case as well.

Proposition 8.1.4. *For any $B \subseteq A$, the augmented 2-core edge set $\mathbf{E}_{2C,B}$ is uniformly random, conditional on the augmented 2-core endpoint set $\mathbf{A}_{2C,B}$.*

Proof. If the 2-core algorithm is modified as described above to find the augmented 2-core, the terminal state will similarly leave the augmented 2-core unexposed, and hence the edge set $\mathbf{E}_{2C,B}$ is again a conditionally uniform random matching of $\mathbf{A}_{2C,B}$ □

If the set B is of size $o(m)$, then we can also show that the size and residual distribution of the augmented 2-core will converge to the same w.e.h.p. limits as for the ordinary 2-core.

Proposition 8.1.5. *Assume $M(\mu) > 1$, let $\boldsymbol{\mathcal{G}}(A, V) = \boldsymbol{\mathcal{G}}(\mu)$ and assume also that $M(\mu_{(A,V)}) \to M(\mu)$. For any $B \subseteq A$ such that $|B| = o(m)$, the following hold w.e.h.p.:*

1. $|\mathbf{A}_{2C,W}|/m \to (1 - z_\mu)^2$,
2. $\boldsymbol{\mu}_{2C,W} \to \mu_{2C}$,
3. $M(\boldsymbol{\mu}_{2C,W}) \to M(\mu)$.

Proof. We recall that the 2-core process analyzed in chapter 6 is customizable, in the sense that the endpoints of residual degree 0 chosen deterministically at odd time steps can be chosen in any way we want. In this case, we specify that the endpoints belonging to B are chosen last.

As in chapter 6, we let $\boldsymbol{\tau}$ be the stopping time of the 2-core process, which occurs when there are no remaining endpoints of residual degree 1. The fact that $M(\mu) > 1$ ensures that $z_\mu < 1$, which in turn implies that $\boldsymbol{\tau} = m - \Omega(m)$ w.e.h.p., so there are $\Omega(m)$ endpoints remaining upon termination.

We may accordingly let $\boldsymbol{\tau}_0$ denote the first time that all remaining endpoints of residual degree 1 belong to B. Now, since $B = o(m)$, then the number of endpoints of residual degree 1 at time $\boldsymbol{\tau}_0$ must be $o(m)$. And, since $\boldsymbol{\tau}_0 \leq \boldsymbol{\tau}$ then w.e.h.p., the total number of remaining endpoints is $\Omega(m)$.

It follows that the residual distribution at time $\boldsymbol{\tau}_0$ must satisfy $\boldsymbol{\mu}_{\boldsymbol{\tau}_0}(0) = o(1)$, but as shown in the proof of Theorem 6.1.1 in chapter 6, for any $\xi < 1 - (1 - z_\mu)^2$, we must have $\boldsymbol{\mu}_t(0) = \Omega(t)$ w.e.h.p. for all $t/m < \chi$. It follows that $\boldsymbol{\tau}_0/m \to 1 - (1 - z_\mu)^2$ w.e.h.p. as well, and hence $|\boldsymbol{\tau} - \boldsymbol{\tau}_0| = o(m)$ w.e.h.p.

The three claims of this proposition are immediate consequences of this fact, since the addition of $o(m)$ endpoints to the 2-core has no asymptotic effect on either the residual distribution or its the first moment. □

8.1.4 BFS in the 2-core

We shall exploit the relationship between the augmented 2-core and shortest paths by restricting our BFS accordingly. First, let us denote the entire 2-core (and not just the endpoint set) of $\mathbf{G}$ by $\mathbf{G}_2C$, and we similarly denote the augmented 2-core.

For any $v \in V$, we may naturally define the v-augmented 2-core using the endpoint set $A(v)$, so $\mathbf{G}_{2C,v} = \mathbf{G}_{2C,A(v)}$. Accordingly,

$$\mathbf{R}_{2C,t}(v)$$

denote the t'th endpoint perimeter, when the BFS is performed in the v-augmented 2-core $\mathbf{G}_{2C,v}$, and we similarly let

$$\boldsymbol{\tau}_{2C}(v, r) = \inf\{t : \mathbf{r}_{2C,t}(v) \geq r\}.$$

An immediate consequence of proposition 8.1.3 is that

$$\mathbf{N}_{2C,t}(v) = \mathbf{N}_t(v) \cap \mathbf{A}_{2C,v}(v),$$

since any path from $A(v)$ to $a \in \mathbf{N}_{2C,t}(v)$ will be entirely contained in $\mathbf{A}_{2C,v}(v)$. The same evidently holds for endpoint perimeters as well, so we have the following corollary.

Corollary 8.1.6. *For any vertex v:*

1. $\mathbf{R}_{2C,t}(v) = \mathbf{R}_t(v) \cap \mathbf{A}_{2C,v}(v)$;
2. $\mathbf{r}_t(v) \geq \mathbf{r}_{2C,t}(v)$;
3. $\boldsymbol{\tau}(v, r) \leq \boldsymbol{\tau}_{2C}(v, r)$;

Proof. Immediate. □

Accordingly, from proposition 8.1.1, we may deduce

$$\boldsymbol{\delta}(u, v) \leq \boldsymbol{\tau}_{2C}(u, \chi) + \boldsymbol{\tau}_{2C}(v, \chi) + 1. \tag{8.4}$$

At first glance, this equation appears to suffer from the same problems as (8.2): if $\boldsymbol{\tau}_{2C}(v, \chi) = \infty$, then it provides no useful information, and even if $\boldsymbol{\tau}_{2C}(v, \chi) < \infty$, we can still only derive an upper bound on distances.

However, restricting BFS to the augmented 2-core ultimately simplifies the analysis, because each vertex we encounter has degree at least 2, or equivalently, every endpoint has residual degree at least 1. Intuitively, this means there are no "dead ends," since if we "enter" such a vertex through one of its endpoints, the other endpoint serves as en "exit." It follows that perimeter sizes during 2-core BFS are generally non-decreasing, and this property facilitates the analysis of the neighborhood growth, particularly for small neighborhoods.

In fact as we shall see, with polynomially high probability, the condition $\boldsymbol{\tau}_{2C}(v, \chi) = \infty$ is equivalent to $\boldsymbol{\tau}_{2C}(v, 3) = \infty$, which implies that as soon as a 2-core perimeter of size

3 is found, (8.4) is applicable. Moreover, again w.p.h.p., $\boldsymbol{\tau}_{2C}(v,3) = \infty$ implies that v does not belong to the giant component of $\mathbf{G}$, and thus we are able to deal such vertices in a qualitatively different manner.

8.2 Theorem Statement and Proof Overview

Our main theorem is regarding for the diameter of the random graph $\boldsymbol{\mathcal{G}}(\mu)$ is the following.

Theorem 8.2.1. *Assume μ satisfies $M(\mu) > 1$ and $\mu(0) > 0$, and let $\mathbf{G} = \boldsymbol{\mathcal{G}}(\mu)$, with the additional assumption that $M(\mu_{\mathbf{G}}) \to M(\mu)$.*

Then the diameter satisfies $\frac{\Delta(\mathbf{G})}{\ln m} \to \Delta_\mu$ w.p.h.p., where

$$\Delta_\mu = \frac{2}{-\ln \psi'_\mu(z_\mu)} + \frac{1}{\ln M(\mu)}, \tag{8.5}$$

where ψ_μ is the p.g.f., and z_μ is the smallest fixed point in $[0,1]$.

We make some brief remarks on this theorem statement.

1. Since the number of non-isolated vertices in $\boldsymbol{\mathcal{G}}(\mu)$ is $\Theta(m)$, then, provided that the number of isolated vertices is $O(m)$, we have $\ln m = \ln n \pm o(1)$, and hence $\ln n$ can be substituted for $\ln m$ in the formula for $\Delta(\mathbf{G})$, and the same convergence $\frac{\Delta(\mathbf{G})}{\ln n} \to \Delta_\mu$ will also hold w.p.h.p.

2. We have imposed the additional assumption of convergence $M(\mu_{\mathbf{G}}) \to M(\mu)$ of the first moment. Of course, if $M(\mu) = \infty$, this will occur anyway, so this is only relevant if $M(\mu) < \infty$.

3. The assumption $M(\mu) > 1$ ensures that the graph contains a giant connected component, while the assumption $\mu(0) > 0$ ensures that the minimum residual degree is 0, or equivalently, the minimum (positive) vertex degree is 1. If $\mu(0) = 0$, the behavior differs qualitatively, and this case is discussed in section 8.8.

4. As discussed in the previous chapter, the value $\psi'_\mu(z_\mu)$ is equal to both $\mu_{2C}(1)$, where μ_{2C} is the 2-core residual distribution, and $M(\mu_{Small})$, where μ_{Small} is the residual distribution of the non-giant components of $\boldsymbol{\mathcal{G}}(\mu)$.

8.2.1 Growth Rates and Large Deviations

We now give some intuitive justification for the formula in (8.5). The first thing that perhaps needs justification is the fact that distances in a random graph are typically of order $\Theta(\ln m) = \Theta(\ln n)$. This is because perimeters grow at an exponential rate, and so if $\mathbf{r}_t(v) \simeq C^t$ for some $C > 1$, it will take $\Theta(\ln m)$ iterations until $\mathbf{r}_t(v)$ reaches size $m^{\Theta(1)}$.

As discussed in the previous chapter, BFS in a random graph loosely resembles a branching process based on the residual distribution μ, in that successive perimeter sizes approximately satisfy

$$\mathbf{r}_{t+1}(v) \simeq \sum_{i=1}^{\mathbf{r}_t(v)} \mathbf{x}_{\mu,i},$$

where the $\mathbf{x}_{\mu,i}$ are i.i.d. μ-distributed. The expected growth rate can thus be approximated by $\mathbf{r}_t(v) \simeq M(\mu)^t \deg(v)$.

This average growth rate yields a rough approximation of the "average" distance between vertices $\boldsymbol{\mathcal{G}}(\mu)$. In order to find a vertex u starting at v, "on average" we expect to also find about half of the vertices (or endpoints) in the entire graph. If perimeters indeed grow exponentially a rate of $M(\mu)$, then at some time around $t = \frac{\ln m}{\ln M(\mu)}$, the perimeter size should be of the order

$$M(\mu)^{\frac{\ln m}{\ln M(\mu)}} = n,$$

assuming of course that $M(\mu) > 1$.

Our strategy for computing the diameter, as described above, involves performing BFS simultaneously from two different vertices until perimeters of size roughly $m^{1/2}$ are found around each vertex. However, this does not affect the above estimate, since if average growth occurs on both sides, then each BFS process will last $t = \frac{\ln m}{2 \ln M(\mu)}$ iterations, for a total distance of $\frac{\ln m}{2 \ln M(\mu)}$.

Although this estimate is fairly informal, it is not too difficult to make the above argument sufficiently rigorous to show the following:

1. that the "average distance" is indeed $\Theta(\ln m)$;
2. that the diameter is also $\Theta(\ln m)$;
3. that the "average distance" is $\frac{\ln m}{\ln M(\mu)} \pm o(\ln m)$.

These are roughly ordered by technical difficulty, but the route to any of these results is reasonably straightforward.

On the other hand, computing the diameter with asymptotic precision is more complicated, since, unlike the average distance, the diameter reflects abnormally slow growth rather than typical behavior of a BFS process. In most cases, the maximum distance will exceed the average distance by $\Theta(\ln n)$, and the most difficult part of the diameter result is to determine exactly how much the diameter exceeds the average distance. Indeed, while it is easy to arrive at the estimate of $\frac{\ln m}{\ln M(\mu)}$ for the average distance, even "guessing" at how much the diameter "should" exceed the average distance is not entirely trivial.

Returning to the formula for the diameter in Theorem 8.2.1, we see that the average distance estimate computed above is represented as the second term in (8.5). Hence, the excess path length due to abnormally slow growth accounts for the first term $\frac{2}{-\ln \psi'(z_\mu)}$. The longest finite distance will generally occur between two vertices which exhibit abnormally slow growth; this slow growth will add a segment of length $\frac{\ln m}{-\ln \psi'_\mu(z_\mu)}$ to both sides of an "average" shortest path. Hence, a more structurally accurate formula for Δ_μ would be

$$\Delta_\mu = \frac{1}{-\ln \psi'_\mu(z_\mu)} + \frac{1}{\ln M(\mu)} + \frac{1}{-\ln \psi'_\mu(z_\mu)}.$$

Large Deviations of a Branching Process

The amount of "excess distance" $\frac{\ln m}{-\ln \psi'_\mu(z_\mu)}$ can be understood in terms of large deviations of the corresponding branching processes. Given a branching process $(\mathbf{y}_t)$ generated by the residual distribution μ, let us denote by $\boldsymbol{\tau}(q)$ the first time that $\mathbf{y}_t$ exceeds q. Of

course, if the process becomes extinct, then we have $\boldsymbol{\tau}(q) = \infty$, but otherwise, if average growth takes place, then we have $\boldsymbol{\tau}(q) \simeq \frac{\ln q}{\ln M(\mu)}$ as described above.

Abnormally slow growth can be characterized in terms of the limit

$$f(c) = \lim_{q\to\infty} \frac{-\ln \mathsf{P}\left[\infty > \boldsymbol{\tau}(q) > \left(c + \frac{1}{\ln M(\mu)}\right) \ln q\right]}{\ln q},$$

and in the nomenclature of large deviations, the function f is called a *rate function*. If such a relationship can be established, then the probability experiencing slow BFS growth in the form of an excess distance of $\frac{c}{\ln m}$ would be $m^{-f(c)}$. In particular, if $f(c) > 1$, then this probability is $m^{-1-\Omega(1)}$, and hence this event does not occur for any of the $n = \Theta(m)$ vertices. Similarly, if $f(c) < 1$ then the expected number of vertices which exhibit such a large deviation would be $n^{\Omega(1)}$, and in this case we expect to find "many" such vertices.

For an actual branching process, rather than a BFS process in a random graph, it is not difficult to compute such a rate function; the reader familiar with branching processes will recognize the significance of the derivative $\psi'_\mu(z_\mu)$ in this context. Perhaps the most expeditions way to compute the above rate function function is to recall that the p.g.f. for the t'th generation of a branching process is the t'th iterate of the original p.g.f.

In this case, as t grows large, the value of $\psi_t(z)$ will tend to the fixed point z_μ for any initial value of $z \in [0, 1]$:

$$\lim_{t\to\infty} \psi_t(z) = \underbrace{\psi \circ \ldots \circ \psi}_{t}(z) \to \psi(z_\mu) = z_\mu.$$

Accordingly, the rate at which $\psi_t(z) \to z_\mu$ is determined by the derivative $\psi'(z_\mu)$, and thus $\psi'(z_\mu)$ provides a first-order estimate of the behavior of $\psi_t(z)$, which can then be used to compute the above rate function.

The 2-core and the Skeleton

Unfortunately, BFS is not a true branching process, and while this direct algebraic computation might be feasible, we shall adopt a different approach which yields a more intuitive interpretation of the corresponding rate function. For this we recall that the

derivative $\psi'(z_\mu)$ is also equal to $\mu_{2C}(1)$, which is the fraction of endpoints of residual degree 1 in the 2-core.

In the previous chapter, we drew a connection between the 2-core of a random graph and the *skeleton* of a Galton-Watson tree, which is the sub-tree consisting all vertices which are themselves the root of an infinite sub-tree. In particular, the skeleton is also a GW tree, and the corresponding distribution is the same as for the 2-core.

Also, any vertex which does not belong to the skeleton of a GW tree quickly becomes extinct, and hence the large deviation rate for a surviving branching process will be the same as for the skeleton process. However, the rate is much easier to estimate for the skeleton, since generation sizes are non-decreasing in the skeleton.

It can be shown rather easily that the bulk of any large deviation will occur at the very beginning of the skeleton process, when the population size is still equal to 1. In this case, the probability of experiencing no growth for t consecutive generations is precisely

$$\mu_{2C}(1)^t = \psi'(z_\mu)^t.$$

As soon as the population size increases to 2, it becomes "twice as difficult" to retard the growth rate in this way. Hence, the "easiest" way to experience a large deviation rate of c is to have no growth for $c \ln q$ generations, and then average growth thereafter.

The probability of such a deviation is $\psi'(z_\mu)^{c \ln q}$, from which we may compute the rate function

$$f(c) = -c \ln \psi'(z_\mu).$$

To estimate of the maximum excess path length in a random graph, we solve for $f(c) = 1$, which leads to the formula $c = \frac{1}{-\ln \psi'(z_\mu)}$.

8.2.2 Upper Bound Proof Overview

The upper bound proof of theorem 8.2.1, which appears in sections 8.3 through 8.6, is more technical than the lower bound proof. In order to achieve this upper bound, we must show that, for any $\epsilon > 0$, the maximum finite distance in $\mathbf{G}$ is w.p.h.p. at most $(\Delta_\mu + \epsilon) \ln m$.

The basic idea is to make rigorous the large deviations argument outlined in §8.2.1, and in so doing prove that, for any pair of vertices u, v, the probability that $\boldsymbol{\delta}(u, v)$ exceeds $(\Delta_\mu + \epsilon) \ln m$ is $m^{-2-\Omega(1)}$, which will imply that w.p.h.p. no such pair of vertices exists.

In order to accomplish this, we must overcome the minor differences between a BFS process and a true branching process which were described in the previous chapter. One of these differences is that the residual distribution does not remain constant during BFS, and in section 8.3, we derive a set of preliminary technical lemmas, which are designed to deal with issues related to the slight changes in residual distribution.

The second source of technical difficulties is the fact that *cross-edges* can occur during BFS:

- a *horizontal edge* is an edge connecting two endpoints in $\mathbf{R}_t(v)$;
- a set of *diagonal edges* is a set of edges which connect two or more endpoints in $\mathbf{R}_t(v)$ to the same vertex.

In section 8.4, we deal individually with both kinds of cross-edges, and the result is a lower bound on the rate of perimeter growth during BFS.

This lower bound will establish a large deviations inequality of the form $\mathsf{P}_t[\mathbf{r}_{t+1}(v)/\mathbf{r}_t(v) < M] < e^{-C\mathbf{r}_t(v)}$ for any growth rate M which is less than the average residual degree $M(\mu)$, and what is essential is that the constant C can be chosen uniformly for the entire BFS process (until the neighborhood size reaches $\chi = m^{1/2} \ln^2 m$). Since faster perimeter growth means less iterations until a "large" perimeter size is reached, then a lower bound on this growth rate is appropriate for achieving an upper bound on distances and the diameter.

However, this large deviations inequality is not useful for cases when $\mathbf{r}_t(v) = O(1)$, and to deal with this situation we exploit the relationship between shortest paths and the 2-core described in §8.1.4. In section 8.5, we analyze the augmented 2-core BFS process, and, using the fact that perimeter sizes in the 2-core are generally non-decreasing, we are able to compute an upper bound in probability on the number of iterations until either the neighborhood size reaches χ or drops to 0.

Finally, in section 8.6, we complete the upper bound proof of theorem 8.2.1. By corollary 8.1.6, analysis of 2-core BFS in section 8.5 will yield an upper bound on the maximum distance between any pair of vertices in the giant component. In order to complete the upper bound proof, we must also consider the distance between vertices which do not belong to the giant component, but it is relatively easy to show the diameter of these small components is w.p.h.p. strictly smaller than the diameter of the giant component.

8.2.3 Lower Bound Proof Overview

The lower bound proof of theorem 8.2.1, which appears in section 8.6, is not quite as technical as the lower bound proof, due largely to the relationship between shortest paths and the 2-core.

In order to bound the diameter from below, it is necessary to show that w.p.h.p. there exists a pair of vertices u, v which are separated by a distance of at least $(\Delta_\mu - \epsilon) \ln m$, for any fixed $\epsilon > 0$. Our strategy is to show that there must exist at least two vertices which are connected by a path to the 2-core, but such that this path has length at least $(\frac{1}{-\ln \psi'(z_\mu)} - \epsilon/2) \ln m$. Once this is established, it only remains to prove that segment of the shortest path connecting u, v which crosses the 2-core has is not significantly shorter than the average distance $\frac{\ln m}{M(\mu)}$, since in this case the total path length will exceed $(\Delta_\mu - \epsilon) \ln m$.

8.3 Technical Preliminaries

Ideally, we would like to reason about the local structure of a random graph as if it were a GW tree, as described in the previous section. Unfortunately, BFS on a random graph is not in fact a proper branching process, and a significant portion of the technical work in this chapter is devoted to overcoming the slight differences between the two, namely the slight dependencies between the residual degrees of endpoints explored during BFS, along with the fact that it is possible, however unlikely, to find a cycle in the local structure of a random graph.

Our strategy for dealing with these "imperfections," is to construct rigorous compar-

isons with "perfect" and well-understood random structures such as i.i.d. random variables and sequences of Bernoulli trials, and in this section, we develop some basic tools to accomplish this.

8.3.1 Sensitivity of the Residual Distribution

We begin with a technical proposition which states that, under our asymptotic assumptions, a vertex set of size $o(m)$ can only contain $o(m)$ endpoints.

Proposition 8.3.1. *For any fixed μ and $\epsilon > 0$ there exists $\delta > 0$ such that the following holds.*

For any endpoint partition (A, V) satisfying $\left\|\mu_{(A,V)} - \mu\right\|_1 < \delta$, and any subset $B \subseteq A$ satisfying $|(B, V)| > (1 - \delta)|A|$, we have

$$|A(V(B))| < \epsilon |A| .$$

Proof. This proposition is more or less an immediate consequence of the uniform summability of the residual distribution with respect to endpoint removal, which was established in theorem 5.1.1 from chapter 5. Nevertheless, for completeness, we give an independent (and brief) proof in terms of explicit constants.

First, for any ϵ_0, we may choose J sufficiently large such that $\sum_{i>J} \mu(i) < \epsilon_0$, and hence by choosing δ sufficiently small, we can guarantee that

$$\frac{|\{a \in A : \mathrm{res}(a) > J\}|}{m} = \sum_{i>J} \mu_{(A,V)}(i) < \epsilon_0$$

where $m = |A|$. It follows that $A(V(B))$ can contain at most $\epsilon_0 m$ endpoints of residual degree greater than J.

Moreover, any $a \in B$ of residual degree J or less can only contribute $J + 1$ total endpoints to $A(V(B))$, so the the number of endpoints in $(A(V(B))$ of residual degree $J+1$ or less is at most $(J + 1)\delta m$, and combining these two allows us to bound $|A(V(B))| < \delta m$. □

Corollary 8.3.2. *For any fixed μ and $\epsilon > 0$ there exists $\delta > 0$ such that the following holds.*

For any endpoint partition (A, V) satisfying $\left\|\mu_{(A,V)} - \mu\right\|_1 < \delta$, and any subset $B \subseteq A$ satisfying $|B| > (1-\delta)|A|$, we have

$$\left\|\mu - \mu_{(V,B)}\right\|_1 < \epsilon$$

Proof. By the above proposition, for arbitrary $\epsilon > 0$, $A(V(B))$ can be made to contain at most $\epsilon_0 m$ endpoints, and hence at least $(1-\epsilon_0)m$ endpoints retain their same residual degree in both (A, V) and (B, V), and we can therefore ensure that $\left\|\mu_{(A,V)} - \mu_{(B,V)}\right\|_1 < \epsilon$, and the claim follows by the triangle inequality. □

8.3.2 Domination of Random Variables

Given distributions μ, ν on $\mathbb{Z}$, we say μ *dominates* ν and write

$$\mu \rhd \nu$$

if $\mathsf{P}[\mathbf{x}_\mu \geq i] \geq \mathsf{P}[\mathbf{x}_\nu \geq i]$ for all i, where $\mathbf{x}_\mu$ and $\mathbf{x}_\nu$ are μ- and ν-distributed random variables, respectively. Accordingly, for random variables $\mathbf{x}, \mathbf{y}$, we write $\mathbf{x} \overset{d}{\rhd} \mathbf{y}$ if $\mathfrak{D}[\mathbf{x}] \rhd \mathfrak{D}[\mathbf{y}]$ and say $\mathbf{x}$ *dominates* $\mathbf{y}$ *in distribution.*

We will use domination as a basic tool to derive lower bounds on large deviation probabilities by comparing an "imperfect" random variable $\mathbf{x}$ to a well-understood random variable $\mathbf{y}$. In this case, if we can prove that $\mathbf{x} \rhd \mathbf{y}$, then any lower bound of the form $\mathsf{P}[\mathbf{y} \geq c] > 1 - \epsilon$ immediately transfers to the random variable $\mathbf{x}$.

Domination can also be expressed in terms of the cumulative distribution function

$$F_\mu(i) = \mathsf{P}[\mathbf{x}_\mu \leq i] = \sum_{j \leq i} \mu(i),$$

since we have, by definition, $\mu \rhd \nu$ if and only if $F_\mu(i) \leq F_\nu(i)$ for all i. Also, note that $\mu \rhd \nu$ if and only if the random variables $\mathbf{x}_\mu$ and $\mathbf{x}_\nu$ can be coupled in such a way that that $\mathsf{P}[\mathbf{x}_\mu \geq \mathbf{x}_\nu] = 1$.

Domination of Sums of Random Variables

It is easy to see that if $\mu_1 \rhd \nu_1$ and $\mu_2 \rhd \nu_2$, then

$$\mathbf{x}_{\mu_1} + \mathbf{x}_{\mu_2} \overset{d}{\rhd} \mathbf{x}_{\nu_1} + \mathbf{x}_{\nu_2},$$

where the above random variables are all independent. The following proposition generalizes this observation.

Proposition 8.3.3. *Let $(\mathbf{y}_0, \ldots, \mathbf{y}_T)$ be a random process with $\mathbf{y}_0 = 0$, and let μ be a distribution such that $\mathfrak{D}_t[\Delta \mathbf{y}_{t+1}] \rhd \mu$ always for all $0 \leq t \leq T-1$. Then $\mathbf{y}_T$ dominates in distribution the sum of T i.i.d. μ-distributed random variables.*

Proof. By induction, it suffices to prove that for a pair of random variables $\mathbf{y}_1, \mathbf{y}_2$, and for distributions μ_1, μ_2, if that $\mathbf{y}_1 \rhd \mu_1$, and $\mathfrak{D}[\mathbf{y}_2 \mid \mathbf{y}_1] \rhd \mu_2$ always then

$$\mathbf{y}_1 + \mathbf{y}_2 \overset{d}{\rhd} \mathbf{x}_{\mu_1} + \mathbf{x}_{\mu_2}$$

where as usual $\mathbf{x}_{\mu_1}$ and $\mathbf{x}_{\mu_2}$ are independent.

In this case, for any y such that $\mathsf{P}[\mathbf{y}_1 = y] > 0$, and for any s, we have

$$\mathsf{P}[\mathbf{y}_1 + \mathbf{y}_2 > s \mid \mathbf{y}_1 = y] \geq \mathsf{P}[\mathbf{x}_{\mu_2} > s - y] = 1 - F_{\mu_2}(s - y),$$

where as above F_{μ_2} denotes the cumulative distribution function. Hence

$$\mathsf{P}[\mathbf{y}_1 + \mathbf{y}_2 > s] = \mathsf{E}\left[\mathsf{P}[\mathbf{y}_1 + \mathbf{y}_2 > s \mid \mathbf{y}_1]\right] \geq \mathsf{E}[1 - F_{\mu_2}(s - \mathbf{y}_1)],$$

and since F_{μ_2} is non-decreasing, then

$$\mathsf{P}[\mathbf{y}_1 + \mathbf{y}_2 > s] \geq \mathsf{E}[1 - F_{\mu_2}(s - \mathbf{x}_{\mu_1})] = \mathsf{P}[\mathbf{x}_{\mu_1} + \mathbf{x}_{\mu_2} > s].$$

□

Domination and Bernoulli Trials

Recall that a *Bernoulli trial* is simply a random variable which takes values in the discrete set $\{0, 1\}$. The distribution of a Bernoulli random variable can thus be described by a single parameter $p = \mathsf{P}[\mathbf{z} = 1]$. Also, note that the sum of t i.i.d. Bernoulli random variables is binomially distributed; we shall denote the binomial distribution with parameters t, p by

$$\mathrm{Bin}_{(t,p)}(i) = \binom{t}{i} p^i (1-p)^{t-i}.$$

Now, proposition 8.3.3 immediately implies that, for any sequence of Bernoulli trials satisfying $\mathsf{P}_t[\mathbf{z}_{t+1}] \geq p$ always for all t, if $\mathbf{Z}_t = \sum_{i=1}^{t} \mathbf{z}_i$, then

$$\mathfrak{D}\left[\mathbf{Z}_t\right] \rhd \mathrm{Bin}_{(t,p)}$$

for all t.

We will often encounter cases in which a bound on the conditional probability $\mathsf{P}_t[\mathbf{x}_{t+1}]$ only holds for $t < \boldsymbol{\tau}$, where $\boldsymbol{\tau}$ is an optional time. The following proposition allows us to argue similarly to above in this slightly trickier situation.

Proposition 8.3.4. *Consider a sequence of Bernoulli random variables $\mathbf{z}_1, \mathbf{z}_2, \ldots$ and let $\mathbf{Z}_t = \sum_{i=1}^{t} \mathbf{z}_i$. Let $\boldsymbol{\tau}$ be an optional time such that, for some $0 < p < 1$, we have*

$$\mathsf{P}_t\left[\mathbf{z}_{t+1} = 1 \mid \boldsymbol{\tau} > t\right] \geq p$$

always for all $t \geq 0$. Then for any $t, r \geq 0$, we have

$$\mathsf{P}[\boldsymbol{\tau} > t \wedge \mathbf{Z}_t \leq r] \leq \mathsf{P}[\mathbf{x}_{\mathrm{Bin}_{(t,p)}} \leq r].$$

Proof. We define a second sequence of Bernoulli trials

$$\mathbf{y}_t = \begin{cases} \mathbf{z}_t \text{ if } t \leq \boldsymbol{\tau} \\ 1 \text{ if } t > \boldsymbol{\tau}, \end{cases}$$

and we let $\mathbf{Y}_t = \sum_{i=1}^{t} \mathbf{y}_t$. Clearly, $\mathsf{P}_t\left[\mathbf{y}_{t+1} = 1\right] \geq p$ always, and thus by proposition 8.3.3, $\mathbf{Y}_t \rhd \mathrm{Bin}_{(t,p)}$ for all t. Moreover, since $\mathbf{Y}_t = \mathbf{Z}_t$ for $t \leq \boldsymbol{\tau}$, then

$$\mathsf{P}[\boldsymbol{\tau} > t \wedge \mathbf{Z}_t \leq r] \leq \mathsf{P}[\mathbf{Y}_t \leq r] \leq \mathsf{P}[\mathbf{x}_{\mathrm{Bin}_{(t,p)}} \leq r].$$

□

8.3.3 The Truncated Distribution

Our methods for asymptotic parametrization guarantee the residual distribution of $\mathbf{G} = \boldsymbol{\mathcal{G}}(A, V)$ converges to a fixed limit μ; this convergence assures us that $\left\|\mu_{(A,V)} - \mu\right\|_1 < \epsilon$ for every fixed $\epsilon > 0$, but we do not have precise knowledge of the residual distribution $\mu_{(A,V)}$. Our arguments must therefore be based on properties of the residual distribution which hold on an entire ϵ-neighborhood of μ (or, equivalently, properties which are continuous with respect to the ℓ_1 norm).

In order to overcome this lack of precise knowledge about the residual distribution, we introduce the notion of a *truncated distribution*. Intuitively, for any $\epsilon > 0$, the ϵ-truncated distribution, which we shall denote by $\mu_{[\epsilon]}$ is the lower bound for the ϵ-neighborhood of μ with respect to the partial order of domination defined above. Now, although the residual distribution has domain $\mathbb{Z}^* = \{0, 1, \ldots\}$, for technical reasons, in this chapter we will extend the domain to include the value $\{-1\}$; the purpose of this will become clear in the next section.

The following proposition establishes the existence of the ϵ-truncated distribution, and the proof shows how to construct $\mu_{[\epsilon]}$ explicitly.

Proposition 8.3.5. *For any distribution μ on $\mathbb{Z}^* \cup \{-1\}$ and any $\epsilon > 0$, there exists a unique distribution $\mu_{[\epsilon]}$ on $\mathbb{Z}^* \cup \{-1\}$ satisfying the following properties:*

1. $\left\|\mu - \mu_{[\epsilon]}\right\|_1 \leq \epsilon$, *and*
2. *every distribution ν on $\mathbb{Z}^* \cup \{-1\}$ with $\|\mu - \nu\|_1 \leq \epsilon$ dominates $\mu_{[\epsilon]}$ (i.e. $\mu_{[\epsilon]} \lhd \nu$).*

Proof. We define the truncated distribution in by its distribution function

$$F_{\mu_{[\epsilon]}}(j) = \max\{F_\mu(j) + \epsilon/2, 1\}$$

for all $j \in \mathbb{Z}^* \cup \{-1\}$.

It is easy to verify that $\left\|\mu - \mu_{[\epsilon]}\right\|_1 \leq \epsilon$. And, for any pair of distributions μ, ν and any j,

$$\|\mu - \nu\|_1 = \sum_{i=-1}^{j} |\mu(i) - \nu(i)| + \sum_{i=j+1}^{\infty} |\mu(i) - \nu(i)| \leq 2\left|F_\nu(j) - F_\mu(j)\right|.$$

If follows that if $\|\mu - \nu\|_1 \leq \epsilon$, then $F_\nu(j) \leq F_\mu(j) + \epsilon/2$ for all j, and hence $\nu \rhd \mu_{[\epsilon]}$. □

Intuitively, then, the truncated distribution $\mu_{[\epsilon]}$ is created by taking an amount $\epsilon/2$ of probability weight from the top of the distribution, and and placing this weight on the value -1.

Domination and I.I.D. Sampling

Our final proposition in this section deals with the sum of $\mathbf{y}$ i.i.d. random variables, where $\mathbf{y}$ is itself a random variable whose distribution dominates a binomial distribution.

Proposition 8.3.6. *Let μ be a distribution on $\mathbb{Z}^* \cup \{-1\}$, let $r \in \mathbb{Z}^*$ and $0 \leq \epsilon \leq 1$, and let $\mathbf{y}$ be a $\mathbb{Z}$-valued random variable such $0 \leq \mathbf{y} \leq r$ always and such that $\mathbf{y} \overset{d}{\rhd} \mathbf{x}_{\mathrm{Bin}_{(r,1-\epsilon)}}$. Then*

$$\sum_{i=1}^{\mathbf{y}} \mathbf{x}_{\mu,i} \overset{d}{\rhd} \sum_{j=1}^{r} \mathbf{x}_{\mu_{[2\epsilon]},j}.$$

Proof. We shall in fact prove the stronger statement

$$\sum_{i=1}^{\mathbf{y}} (\mathbf{x}_{\mu,i} + 1) \overset{d}{\rhd} \sum_{j=1}^{r} (\mathbf{x}_{\mu_{[2\epsilon]},j} + 1),$$

Note that since $\mathbf{x}_\mu + 1 \geq 0$ always, and since $\mathbf{y} \overset{d}{\rhd} \mathbf{x}_{\mathrm{Bin}_{(r,1-\epsilon)}}$, we have

$$\sum_{i=1}^{\mathbf{y}} (\mathbf{x}_{\mu,i} + 1) \overset{d}{\rhd} \sum_{j=1}^{\mathbf{x}_{\mathrm{Bin}_{(r,1-\epsilon)}}} (\mathbf{x}_{\mu,j} + 1).$$

It therefore suffices to prove

$$\sum_{i=1}^{\mathbf{x}_{\mathrm{Bin}_{(r,1-\epsilon)}}} (\mathbf{x}_{\mu,i} + 1) \overset{d}{\rhd} \sum_{j=1}^{r} (\mathbf{x}_{\mu_{[2\epsilon]},j} + 1). \tag{8.6}$$

Now, let $\mathbf{I}_{1-\epsilon}$ denote a Bernoulli random variable with $\mathsf{P}[\mathbf{I}_{1-\epsilon} = 1] = 1 - \epsilon$ and $\mathsf{P}[\mathbf{I}_{1-\epsilon} = 0] = \epsilon$. Since $\mathbf{x}_{\mathrm{Bin}_{(r,1-\epsilon)}} \stackrel{d}{=} \sum_{j=1}^{r} \mathbf{I}_{1-\epsilon,j}$, where as usual the $\mathbf{I}_{1-\epsilon,j}$ are independent, we can express the left hand side of the above expression as

$$\sum_{i=1}^{\mathbf{x}_{\mathrm{Bin}_{(r,1-\epsilon)}}} (\mathbf{x}_{\mu,i} + 1) \stackrel{d}{=} \sum_{j=1}^{r} \mathbf{I}_{1-\epsilon,j}(\mathbf{x}_{\mu,j} + 1). \tag{8.7}$$

It is clear that

$$\left\| \mathfrak{D}\left[\mathbf{I}_{1-\epsilon}(\mathbf{x}_\mu + 1)\right] - \mathfrak{D}\left[\mathbf{x}_\mu + 1\right] \right\|_1 \leq 2\epsilon,$$

and therefore

$$\mathbf{I}_{1-\epsilon}(\mathbf{x}_\mu + 1) \stackrel{d}{\rhd} \mathbf{x}_{\mu_{[2\epsilon]}} + 1 \tag{8.8}$$

by proposition 8.3.5. Equation (8.6) now follows from (8.8) and (8.7), and the proof is complete. □

8.4 Lower Bound for Growth of Endpoint Perimeters

In this section we derive a large deviations lower bound for perimeter growth during BFS. The theorem statement is as follows.

Theorem 8.4.1. *For any fixed distribution μ and any $M < M(\mu)$, there exist $\delta > 0$ and $C > 0$ such that the following statement holds.*

Let $H = (A, V)$ be an endpoint partition satisfying $\|\mu_H - \mu\|_1 < \delta$. For any $v \in V$ and any $t \geq 0$, if $|\mathbf{N}_t(v)| < \delta m$ then

$$\mathsf{P}_t\left[\mathbf{r}_{t+1}(v) < M \cdot \mathbf{r}_t(v)\right] \leq e^{-C \cdot \mathbf{r}_t(v)}. \tag{8.9}$$

What is important is that the same constant C can be used for all iterations of BFS until the total neighborhood size exceeds δm. The proof of this theorem is organized as follows. First, in 8.4.1, we show that it suffices to consider a single iteration of BFS, and establish domination with respect to an arbitrary truncated distribution.

Then, in §8.4.2 and §8.4.3, we deal with the two kinds of cross-edges that cause a BFS process to differ from a true branching process. In §8.4.4, we prove the single iteration

version of Theorem 8.4.1, which is stated below, and this will imply the general theorem statement above.

8.4.1 One Iteration of BFS

Due to the stability conditions on the residual distribution we are able to consider a single iteration of BFS at a time. Recall that a subset $R \subseteq A$ is *vertex-closed* if $R = A(V(R))$, meaning that for any vertex $v \in V$, either all or none of the endpoints in $A(v)$ belong to R. Each endpoint perimeter $\mathbf{R}_t(v)$ is by definition vertex-closed with respect to the endpoint partition $(\mathbf{A}_t(v), V)$, where $\mathbf{A}_t(v) = A \backslash \mathbf{N}_t(v)$, and thus the general setting for an arbitrary BFS iteration can be described in terms of a vertex-closed subset as in the following theorem.

Theorem 8.4.2. *For any fixed distribution μ and any $\epsilon > 0$ there exists a fixed $\delta > 0$ such that the following statement holds.*

Let $H = (A, V)$ be an endpoint partition satisfying $\|\mu_H - \mu\|_1 < \delta$. For any vertex-closed endpoint set R satisfying $|R| = r < \delta m$, the set $\mathbf{R}_1 = A(V(\vec{\mathbf{E}}(R))) \backslash R$ satisfies

$$|\mathbf{R}_1| \rhd \sum_{i=1}^{r} \mathbf{x}_{\mu_{[\epsilon],i}}$$

Assuming we are able to establish this result, the Theorem 8.4.1 can be proved as follows.

Proof of Theorem 8.4.1. During BFS, the perimeter $\mathbf{R}_{t+1}(v)$ is precisely equal to

$$\mathbf{R}_{t+1}(v) = A(V(\vec{E}(\mathbf{R}_t(v)))) \backslash \mathbf{R}_t(v)$$

by definition. Also by definition, the set $\mathbf{R}_t(v)$ is vertex-closed with respect to $(\mathbf{A}_t(v), V)$, and hence theorem 8.4.2 will imply that, for any $\epsilon_0 > 0$, there exists $\delta_0 > 0$ such that

$$\mathfrak{D}_t[\mathbf{r}_{t+1}(v)] \rhd \sum_{i=1}^{\mathbf{r}_t(v)} \mathbf{x}_{\mu_{[\epsilon_0],i}} \tag{8.10}$$

whenever both

1. $\left\|\mu_{(\mathbf{A}_t(v),V)} - \mu\right\|_1 < \delta_0$, and

2. $|\mathbf{R}_t(v)| < \delta_0 \, |\mathbf{A}_t(v)|$.

Since we are able to bound $|\mathbf{N}_t(v)| < \delta \, |A|$, for arbitrary $\delta > 0$, both of these conditions can be satisfied using proposition 8.3.1 and corollary 8.3.2 from §8.3.1. For the second condition, since $\mathbf{R}_t(v) \subseteq A(V(\mathbf{N}_t(v)))$, then $|\mathbf{R}_t(v)|$ can be made arbitrarily small due to proposition 8.3.1, and since $\mathbf{A}_t(v) = A \backslash \mathbf{N}_t(v)$, then the ratio $|\mathbf{R}_t(v)| \, / \, |\mathbf{A}_t(v)|$ can be made arbitrarily small as well. The first condition can similarly be satisfied by due to corollary 8.3.2, again using the fact that $|\mathbf{A}_t| > (1 - \delta) \, |A|$ for arbitrary $\delta > 0$.

To achieve the large deviation bound in (8.9), we first claim that ϵ_0 can be chosen sufficiently small such that $M(\mu_{[\epsilon_0]}) > M$ for any fixed $M < M(\mu)$. Indeed, by definition,

$$M(\mu) = \sum_{i=0}^{\infty} i\mu(i) = \lim_{J \to \infty} \sum_{i=0}^{J} i\mu(i).$$

We may now choose J such that the partial sum $M_J(\mu) = \sum_{i=0}^{J} i\mu(i)$ exceeds M, and since clearly $M(\mu_{[\epsilon_0]}) \geq M_J(\mu) - \epsilon_0 - J \cdot \epsilon_0$, then we can ensure that $M(\mu_{[\epsilon_0]}) > M$.

Finally, due to Chernoff's large deviation inequality (theorem 2.3.1 from chapter 2), the fact that the truncated distribution satisfies $\mathsf{P}[\mathbf{x}_{\mu_{[\epsilon_0]}} < -1] = 0$ ensures that

$$\mathsf{P}\left[\sum_{i=1}^{r} \mathbf{x}_{\mu_{[\epsilon_0]},i} < M\right] < e^{-C \cdot r}$$

for some fixed constant C. □

8.4.2 Horizontal Edges

A *horizontal edge* occurs if two endpoints in $\mathbf{R}_t(v)$ match to each other, and we now derive a bound in probability on the number of horizontal edges which may occur in an iteration of BFS. Intuitively, if we are given an subset $R \subseteq A$ of such that $|R| = r$ and $|A| = m$ and $r \leq m/2$, the probability that any given endpoint ends up in a horizontal edge is $\frac{r-1}{m-1} < r/m$.

In this case, of course, a second endpoint is also involved in this horizontal edge. The following lemma demonstrates that, for the purposes of establishing an upper bound, this probability can be "shared" between both endpoints, in the sense that each endpoint can be considered to have independently been landed in a horizontal edge with probability $\sqrt{r/m}$.

Lemma 8.4.3. *Consider a subset R of an endpoint set A, let $r = |R|$ and $m = |A|$, and assume $r \leq m/2$. For a uniformly random matching $\mathbf{E} = \boldsymbol{\mathcal{E}}(A)$, we have*

$$\mathfrak{D}\left[\left|\vec{\mathbf{E}}(R) \cap R\right|\right] \lhd \mathrm{Bin}_{(r,\sqrt{r/m})}\,. \tag{8.11}$$

Proof. We first define

$$f(s,r,m) = \mathsf{P}\left[\left|\vec{\mathbf{E}}(R) \cap R\right| \geq s\right]$$

for $r = |R|$ and let $m = |A|$, and similarly $g(s,p,r) = \mathsf{P}[\mathbf{x}_{\mathrm{Bin}_{(r,p)}} \geq s]$. Equation (8.11) is therefore equivalent to the statement that

$$f(s,r,m) \leq g\left(s, \sqrt{r/m}, r\right) \tag{8.12}$$

holds for all s.

We proceed by induction; clearly (8.12) holds for $r = 0, 1$ and any s and $m \geq 2r$, since if $|R| \leq 1$ then there cannot be any internally matched endpoints. Now, choose any r and assume inductively that (8.12) holds for all triples (s', r', m') where $r' < r$ and $m' \geq 2r'$.

We note that any given endpoint $a \in R$ matches to another endpoint within R with probability $(r-1)/(m-1)$. If this occurs, then $r-2$ unexposed endpoint remain in R, and we have 2 internally matched endpoints. Otherwise, we have $r-1$ remaining unexposed endpoints in R, and no internally matched endpoint. In both cases, the total number of unexposed endpoints remaining is $m-2$. Hence, we inductively compute

$$\begin{aligned} f(s,r,m) &= \frac{r-1}{m-1} f(s-2, r-2, m-2) + \frac{m-r}{m-1} f(s, r-1, m-2) \\ &\leq \frac{r-1}{m-1} g\left(s-2, \sqrt{\tfrac{r-2}{m-2}}, r-2\right) + \frac{m-r}{m-1} g\left(s, \sqrt{\tfrac{r-1}{m-2}}, r-1\right) \\ &\leq (r/m) g\left(s-2, \sqrt{r/m}, r-2\right) + (1-r/m) g\left(s, \sqrt{r/m}, r-1\right). \end{aligned}$$

The inequality in (8.12) now follows by letting $p = \sqrt{r/m}$ and computing

$$\begin{aligned} g(s,p,r) &= p \cdot g(s-1,p,r-1) + (1-p) \cdot g(s,p,r-1) \\ &= \Big(p^2 \cdot g(s-2,p,r-2) + p(1-p) \cdot g(s-1,p,r-2)\Big) + (1-p) \cdot g(s,p,r-1) \\ &\geq p^2 \cdot g(s-2,p,r-2) + (1-p^2) \cdot g(s,p,r-1). \end{aligned}$$

□

8.4.3 Diagonal Edges and Uniform Endpoint Sampling

We now consider *diagonal edges*, which occur if two endpoints in $\mathbf{R}_{v,i}$ match to the same vertex outside of $\mathbf{N}_{v,i}$. Our approach is based on the following observation. For any integer $r \geq 0$, conditional on the event that exactly r endpoints in $\mathbf{R}_t(v)$ match externally, the set

$$\mathbf{S} = \vec{\mathbf{E}}(\mathbf{R}_t(v)) \backslash \mathbf{R}_t(v),$$

will be a uniformly random subset of size r from $\mathbf{A}_t(v) \backslash \mathbf{R}_t(v)$.

Moreover, the size of the next perimeter in this case will be

$$\mathbf{R}_{t+1}(v) = A(V(\mathbf{S})) \backslash \mathbf{S}.$$

Hence, in the following lemma, we derive a probabilistic lower bound on the size of the set $A(V(\mathbf{S})) \backslash \mathbf{S}$, where $\mathbf{S}$ is a uniform sample of size r from an endpoint partition (A, V).

Lemma 8.4.4. *For any fixed distribution μ and any $\epsilon > 0$, there exists a $\delta > 0$ such that the following statement holds.*

Let (A, V) be an endpoint partition satisfying $\left\|\mu_{(A,V)} - \mu\right\|_1 < \delta$, and consider a uniformly random subset $\mathbf{S} \subseteq A$ of size $|\mathbf{S}| = r < \delta |A|$. Then the random variable $|A(V(\mathbf{S})) \backslash \mathbf{S}|$ dominates in distribution the sum of r i.i.d. $\mu_{[\epsilon]}$-distributed random variables.

Proof. We consider the random process $(\mathbf{s}_1, \ldots, \mathbf{s}_r)$ corresponding to sampling from A without replacement, so at each step t, we choose an endpoint $\mathbf{s}_t$ uniformly at random from $A \backslash \mathbf{S}_{t-1}$, where $\mathbf{S}_t = \{\mathbf{s}_1, \ldots, \mathbf{s}_i\}$ for $t > 0$ and $\mathbf{S}_0 = \emptyset$. We also define $\mathbf{Q}_t = A(V(\mathbf{S}_t)) \backslash \mathbf{S}_t$,

and we write $\mathbf{q}_t = |\mathbf{Q}_t|$. The lemma states that $\mathbf{q}_r$ dominates in distribution the sum of r i.i.d. $\mu_{[\epsilon]}$-distributed random variables.

At time t, we define the *effective residual degree* of any endpoint in $A \backslash \mathbf{S}_t$ by

$$\mathbf{res}_t^*(a) = \begin{cases} -1 & \text{if } a \in \mathbf{Q}_t; \\ \text{res}(a) & \text{otherwise.} \end{cases}$$

We denote the distribution of effective residual degrees at time t by

$$\boldsymbol{\nu}_t(j) = \frac{|\{a \in A \backslash \mathbf{S}_t : \mathbf{res}_t^*(a) = j\}|}{|A \backslash \mathbf{S}_t|},$$

noting that $\mathbf{q}_{t+1} - \mathbf{q}_t = \mathbf{res}_t^*(\mathbf{s}_{t+1})$, and therefore $\mathfrak{D}_t[\mathbf{q}_{t+1} - \mathbf{q}_t] = \boldsymbol{\nu}_t$, where $\mathfrak{D}_t$ is the distribution conditional on the history $(\mathbf{s}_1, \ldots, \mathbf{s}_t)$

We now claim that, for δ sufficiently small, we have $\|\boldsymbol{\nu}_t - \mu\| < \epsilon$ always for all t. Indeed, since $|\mathbf{S}_t| = t \leq r < \delta |A|$, then by corollary 8.3.2, δ can be chosen sufficiently small that, for arbitrary $\epsilon_0 > 0$, the residual distribution at time t satisfies $\left\|\mu_{(A_t,V)} - \mu\right\|_1 < \epsilon_0$ always. And, the actual residual degree and the effective residual degree at time t differ only on the set $\mathbf{Q}_t$; since $\mathbf{Q}_t$ is contained in at most $t \leq r < \delta |A|$ vertices, then by proposition 8.3.1, we can again choose δ sufficiently small such that $|\mathbf{Q}_t| < \epsilon_0 |A|$ always. By setting ϵ_0 appropriately, it follows that δ can be chosen sufficiently small that $\|\boldsymbol{\nu}_t - \mu\| < \epsilon$ always for all t.

It now follows by proposition 8.3.5 that $\boldsymbol{\nu}_t \rhd \mu_{[\epsilon]}$ always, and since $\mathfrak{D}_t[\mathbf{q}_{t+1} - \mathbf{q}_t] = \boldsymbol{\nu}_t$, the proof is complete due to proposition 8.3.3. □

8.4.4 Proof of Theorem 8.4.2

In this section, we combine lemmas 8.4.3 and 8.4.4, to prove theorem theorem 8.4.2, which states that δ can be chosen sufficiently small to ensure that if $\left\|\mu_{(A,V)} - \mu\right\|_1 < \delta$, then for any vertex-closed $R \subseteq A$ of size $|R| = r < \delta m$, we have

$$|\mathbf{R}_1| \lhd \sum_{i=1}^{r} \mathbf{x}_{\mu_{[\epsilon],i}}.$$

Proof of Theorem 8.4.2. First, let $\mathbf{S} = R \backslash \vec{\mathbf{E}}(R)$ denote the subset of R consisting of endpoints which match outside of R, and let $\mathbf{y} = |\mathbf{S}|$. For arbitrary $\epsilon_0 > 0$, we may choose δ sufficiently small that $\sqrt{r/m} < \epsilon_0$, where $m = |A|$, and thus by lemma 8.4.3, we have

$$\mathfrak{D}[\mathbf{y}] \rhd \mathrm{Bin}_{(r,1-\sqrt{r/m})} \rhd \mathrm{Bin}_{(r,1-\epsilon_0)}. \tag{8.13}$$

Next, let $B = A \backslash R$, and by corollary 8.3.2, $\delta > 0$ can be chosen sufficiently small to ensure that $\|\mu_{B,V} - \mu\|_1 < \delta_0$ for arbitrary $\delta_0 > 0$. Conditional on $\mathbf{y}$, the set of endpoints $\vec{\mathbf{E}}(R) \backslash R$ is a uniformly random subset of size $\mathbf{y}$ chosen from B, and thus by by lemma 8.4.4, we have

$$\mathfrak{D}[\mathbf{r}_1 \mid \mathbf{y}] \rhd \mathfrak{D}\left[\sum_{j=1}^{\mathbf{y}} \mathbf{x}_{\mu_{[\epsilon_1]},j}\right] \tag{8.14}$$

for arbitrary $\epsilon_1 > 0$.

Since $\mathbf{y}$ dominates the binomial distribution $\mathrm{Bin}_{(r,1-\epsilon_0)}$, we may invoke proposition 8.3.6 regarding the sum of a random number of random variables, and since $\left(\mu_{[\epsilon_1]}\right)_{[2\epsilon_0]} = \mu_{[2\epsilon_0+\epsilon_1]}$, it follows from (8.13) and (8.14) that $\mathfrak{D}[\mathbf{r}_1] \rhd \mathfrak{D}\left[\sum_{j=1}^{r} \mathbf{x}_{\mu_{[2\epsilon_0+\epsilon_1]},j}\right]$. The proof is now complete since ϵ_0, ϵ_1 can be made arbitrarily small. □

8.5 BFS and the 2-core

In this section, we analyze a BFS process which takes place in a graph with minimum degree 2. In the next section, we will apply these results to the augmented 2-core of an arbitrary random graph $\boldsymbol{\mathcal{G}}(\mu)$, but since the residual distribution of augmented 2-core converges w.e.h.p. to μ_{2C} by proposition 8.1.5, then it suffices to analyze a BFS process in a random graph $\boldsymbol{\mathcal{G}}(\mu_{2C})$ with this same residual distribution, as follows.

Assumptions 8.5.1. In this section, we are considering a BFS process $(\mathbf{r}_t(v))$ in a random graph $\boldsymbol{\mathcal{G}}(A, V)$, under the following conditions:

1. $\mathbf{G} = \boldsymbol{\mathcal{G}}(\mu_{2C})$, where μ_{2C} satisfies both $M(\mu_{2C}) > 1$ and $\mu_{2C}(1) = 0$;
2. every vertex other that (possibly) v has degree at least 2;

3. the process is halted at the first time step $\mathbf{T}$ such that one of the following hold:

   (a) $\mathbf{T} \geq \ln^2 m$,

   (b) $\mathbf{r}_{\mathbf{T}}(v) = 0$,

   (c) $\mathbf{r}_{\mathbf{T}}(v) \geq \chi$, where $\chi = m^{1/2} \ln^2 m$.

We will simplify our notation by omitting the "(v)" from our notation where appropriate. Note that the random time $\mathbf{T}$ is clearly optional, meaning that whether $\mathbf{T} = t$ can be determined from the history at time t. Accordingly, conditional probabilities, expectations, and distributions are accompanied by the assumption that $t < \mathbf{T}$, meaning that $\mathsf{E}_t[\mathbf{r}_{t+1}]$ is in fact the conditional expectation $\mathsf{E}_t[\mathbf{r}_{t+1} \mid t < \mathbf{T}]$.

Recall that $\boldsymbol{\tau}(v, r)$ denotes the first time that $\mathbf{r}_t(v)$ reaches r; we will accordingly abbreviate by $\boldsymbol{\tau}(r)$, and we also define a second hitting time by

$$\boldsymbol{\rho} = \boldsymbol{\rho}(v) = \min\{t : \mathbf{r}_t(v) = 0\}. \tag{8.15}$$

It is clear that, unlike $\boldsymbol{\tau}(v, r)$, the hitting time $\boldsymbol{\rho}(v)$ must be finite.

Recall also that the leading constant for the diameter of $\boldsymbol{\mathcal{G}}(\mu)$ is

$$\Delta_\mu = \frac{2}{-\ln \psi'(z_\mu)} + \frac{1}{\ln M(\mu)} = \frac{2}{-\ln \mu_{2C}(1)} + \frac{1}{M(\mu_{2C})}.$$

Our main result in this section is the following theorem.

Theorem 8.5.1. *For any distribution μ_{2C} and any $\epsilon > 0$, given assumptions 8.5.1:*

1. *with probability $m^{-1-\Omega(1)}$, either $\boldsymbol{\tau}(\chi) < \left(\frac{\Delta_\mu}{2} + \epsilon\right) \ln m$ or $\boldsymbol{\rho} < \epsilon \ln m$;*

2. *$\boldsymbol{\tau}(\chi) < \infty$ with probability at most $m^{-1+o(1)}$.*

The proof of this theorem exploits the assumption the minimum degree in $\mathbf{G}$ is at least 2, and therefore perimeter sizes are generally non-decreasing; this monotonicity property is established in §8.5.1.

Then, in §8.5.2 and §8.5.3, we consider two different phases of the BFS process, and in each one we consider "good" and "bad" iterations. Once the number of good iterations exceeds a certain amount, we are able to conclude that either $\mathbf{r}_t$ must be equal to 0, or else the perimeter has grown by a specified amount. Due to the monotonicity of perimeter sizes, "bad" iterations are not so bad after all: the progress made during "good" iterations cannot be undone, and hence a "bad" iteration is simply a delay rather than a step backward.

Finally, in §8.5.4, we tie up some loose ends related to the small possibility that $\mathbf{r}_t$ never reaches χ, in which case $\boldsymbol{\tau}(\chi) = \infty$.

8.5.1 Monotonicity of Perimeter Sizes

The consequences of the minimum degree assumption for monotonicity of perimeter growth can be summarized as follows.

Lemma 8.5.2. *For any μ_{2C}, and given assumptions 8.5.1, the following hold uniformly for $t < \mathbf{T}$:*

1. $\mathsf{P}_t\left[\mathbf{r}_{t+1} < \mathbf{r}_t\right] = O(\mathbf{r}_t^3/m)$,
2. $\mathsf{P}_t\left[\mathbf{r}_{t+1} < \mathbf{r}_t - 2\right] = O(\mathbf{r}_t^6/m^2)$,
3. $\mathsf{P}\left[\mathbf{r}_{t+1} = \mathbf{r}_t\right] \leq (\mu_{2C}(1) + o(1))^{\mathbf{r}_t} + O(\mathbf{r}_t^3/m)$.
4. *for any constant $M < M(\mu_{2C})$,*

$$\mathsf{P}_t\left[\mathbf{r}_{t+1} < M \cdot \mathbf{r}_t\right] = e^{-\Omega(\mathbf{r}_t)}.$$

Proof. First, note that, since both $\mathbf{r}_t < \chi$ and $t < \ln^2 m$ for all $t < \mathbf{T}$, then clearly $|\mathbf{N}_t(v)| = o(m)$ for the duration of this stopped BFS process, and hence theorem 8.4.1 is applicable with respect to the same constant C for the entire duration. Similarly, the assumption that $\mu_{(A,V)} \to \mu_{2C}$, along with the halting condition, ensure that, for any $\epsilon > 0$, the unexposed residual distribution $\boldsymbol{\mu}_t = \mu_{\mathbf{A}_t}$ satisfies $\|\boldsymbol{\mu}_t - \mu\|_1 < o(1)$ always, and uniformly for all $t < \mathbf{T}$.

Now, for each $a \in \mathbf{R}_t$, define a random variable $\mathbf{y}_{t+1}(a)$ by letting $\mathbf{y}_{t+1}(a) = 1$ if either

1. $\vec{\mathbf{E}}(a) \in \mathbf{R}_t$, or

2. $V(\vec{\mathbf{E}}(a)) = V(\vec{\mathbf{E}}(a')) = u$ for some vertex u with $\deg(u) \leq 2\mathbf{r}_t$,

and $\mathbf{y}_{t+1}(a) = 0$ otherwise. We also define $\mathbf{y}_{t+1} = \sum_{a \in \mathbf{R}_t} \mathbf{y}_{t+1}(a)$, and since every vertex other than v has degree at least 2 by assumption, it follows that

$$\mathbf{r}_{t+1} \geq \mathbf{r}_t - \mathbf{y}_{t+1}.$$

Since $|\mathbf{A}_t| = m - o(m)$ by assumption, then for any given pair $a, a' \in \mathbf{R}_i$, we have $\mathsf{P}_t[\vec{\mathbf{E}}(a) = a'] = O(1/|\mathbf{A}_t|) = O(1/m)$. Also, the probability that a' and a' match to the same vertex u of degree at most $2\mathbf{r}_t$ is $O(\mathbf{r}_t/m)$, and by considering all pairs $a, a' \in \mathbf{R}_t$, it follows that

$$\mathsf{P}_t\left[\mathbf{y} > 0\right] = O\left(\mathbf{r}_t^3/m\right).$$

Similarly, by considering all possible ways that at least 3 endpoints in $\mathbf{R}_t$ can be involved in cross-edges of the kinds described above, we easily deduce

$$\mathsf{P}_t\left[\mathbf{y} > 2\right] = O\left(\mathbf{r}_t^6/m^2\right).$$

For the third statement of the lemma, we note that conditional probability at time t that every endpoint in $\mathbf{R}_t$ matches to an endpoint of residual degree 1 outside of $\mathbf{R}_t$ is at most

$$\boldsymbol{\mu}_t(1)^{\mathbf{r}_t} \leq (\mu_{2C}(1) + o(1))^{\mathbf{r}_t}.$$

And, the event $\mathbf{r}_{t+1} = \mathbf{r}_t$ can only occur if either $\mathbf{y} > 0$ or if every endpoint in $\mathbf{R}_t$ matches to an endpoint of residual degree 1.

The final claim is an immediate implication of theorem 8.4.1. □

Corollary 8.5.3. *The following hold with probability at least* $1 - m^{-2+o(1)}$*:*

1. $\mathbf{r}_{t+1} < \mathbf{r}_t$ *occurs at most once for all* $t < \mathbf{T}$*;*

2. $\mathbf{r}_{t+1} < \mathbf{r}_t - 2$ *never occurs for* $t < \mathbf{T}$*;*

Proof. For the first claim, we may define indicator random variables $\mathbf{z}_t$ such that $\mathbf{z}_t = 1$ if and only if both $\mathbf{r}_t < \mathbf{r}_{t-1}$ and $t \leq \mathbf{T}$, in which case, by the above lemma, we have

$$\mathsf{P}_t[\mathbf{z}_{t+1} = 1] = O(\mathbf{r}_t^3/m).$$

It follows that if $\mathbf{r}_t$ is less than, say, $\ln^2 m$, then $\mathsf{P}_t[\mathbf{z}_{t+1} = 1] = m^{-1+o(1)}$.

Also, since $M(\mu_{2C}) > 1$, then by invoking the fourth claim of the above lemma, we also have

$$\mathsf{P}_t[\mathbf{z}_{t+1} = 1] = e^{-\Omega(\mathbf{r}_t)}.$$

Hence if $\mathbf{r}_t > \ln^2 m$, we also have $\mathsf{P}_t[\mathbf{z}_{t+1} = 1] \leq m^{-1+o(1)}$.

It follows from proposition 8.3.3 that the total number of occurrence of $\mathbf{z}_t$ is dominated by a binomial distribution

$$\mathrm{Bin}_{(\ln^2 m, m^{-1+o(1)})},$$

and thus the probability of at least two occurrences is at most $m^{-2+o(1)} \ln^4 m = m^{-2+o(1)}$.

For the second claim we similarly define $\mathbf{y}_t = 1$ if and only if $\mathbf{r}_t < \mathbf{r}_{t-1} - 2$ and $t \leq \mathbf{T}$, and by the identical argument, $\mathsf{P}_t[\mathbf{y}_{t+1}] \leq m^{-2+o(1)}$, and we may similarly deduce that the probability of one or more occurrences of this event is at most $m^{-2+o(1)}$. □

8.5.2 Phase 1: $0 < \mathbf{r}_t < \ln \ln m$

The first phase of the BFS process lasts until either $\mathbf{r}_t$ reaches $\ln \ln m$ or drops to 0. The value $\ln \ln m$ is somewhat arbitrary; any non-constant function of m which grows slower than $\ln m$ will do. In any case, a "good" iteration in this phase is simply an iteration in which $\mathbf{r}_t > \mathbf{r}_{t-1}$, and it evident by corollary 8.5.3 that there can be at most $\ln \ln m + 2$ good iterations in phase 1.

To account for the possibility that $\mathbf{r}_t$ reaches 0, we define, for any r,

$$\boldsymbol{\rho}(r) = \min\{\boldsymbol{\rho}, \boldsymbol{\tau}(r)\},$$

so $\boldsymbol{\rho}(r)$ is the first time that either $\mathbf{r}_t = 0$ or $\mathbf{r}_t$ exceeds r. Phase 1 thus ends at time $\boldsymbol{\rho}(\ln \ln m)$, and we may bound the duration of this phase as follows.

Lemma 8.5.4. *For any fixed $\epsilon > 0$:*

$$\mathsf{P}\left[\boldsymbol{\rho}(\ln\ln m) > \left(\frac{1}{-\ln \mu_{2C}(1)} + \epsilon\right)\ln m\right] = m^{-1-\Omega(1)}.$$

Proof. For $t \geq 1$, we define indicator random variables

$$\mathbf{z}_t = \begin{cases} 1 & \text{if } \mathbf{r}_t > \mathbf{r}_{t-1} \\ 0 & \text{otherwise,} \end{cases}$$

and we let $\mathbf{Z}_t = \sum_{i=1}^{t} \mathbf{z}_t$. Since $\mathbf{r}_t > \mathbf{r}_{t-1}$ whenever $\mathbf{z}_t = 1$, and since $\mathbf{r}_0 \geq 1$, then corollary 8.5.3 implies that

$$\mathsf{P}[\mathbf{r}_t \geq \mathbf{Z}_t - 1 \text{ for all } t \leq \mathbf{T}] = m^{-2+o(1)},$$

since $\mathbf{r}_t$ can only decrease by at most 2, and this can only occur once.

Since the maximum duration before halting is $\ln^2 m$, for any $t < \ln^2 m$, we have

$$\mathsf{P}\left[\boldsymbol{\rho}(\ln\ln m) > t \wedge \mathbf{Z}_t > \ln\ln m\right] = m^{-2+o(1)}. \tag{8.16}$$

Also, by lemma 8.5.2, we have

$$\mathsf{P}_t\left[\mathbf{z}_{t+1} = 0 \mid \boldsymbol{\rho}(\ln\ln m) > t\right] \leq \mu_{2C}(1) + o(1)$$

always for all $t < \mathbf{T}$. It thus follows from (8.16) and proposition 8.3.4 regarding domination of Bernoulli trials, that

$$\begin{aligned}\mathsf{P}\left[\boldsymbol{\rho}(\ln\ln m) > t\right] &\leq \mathsf{P}\left[\boldsymbol{\rho}(\ln\ln m) > t \wedge \mathbf{Z}_t \leq \ln\ln m\right] + m^{-2+o(1)} \\ &\leq \mathsf{P}\left[\mathbf{x}_{\mathrm{Bin}_{(t,1-\mu_{2C}(1)-o(1))}} \leq \ln\ln m\right] + m^{-2+o(1)}.\end{aligned}$$

We now invoke a simple binomial inequality

$$\mathsf{P}[\mathbf{x}_{\mathrm{Bin}_{(t,1-p)}} \leq r] \leq \binom{t}{r} p^{t-r} \leq \frac{t^r p^t}{p^r},$$

and for $t = c\ln m$, $r = \ln\ln m$, and $p = \mu_{2C}(1) + o(1)$, we deduce

$$\begin{aligned}\mathsf{P}\left[\boldsymbol{\rho}(\ln\ln m) > c\ln m\right] &\leq \left(\frac{c\ln m}{\mu_{2C}(1) + o(1)}\right)^{\ln\ln m} (\mu_{2C}(1) + o(1))^{c\ln m} + m^{-2+o(1)} \\ &\leq m^{c\ln\mu_{2C}(1)+o(1)} + m^{-2+o(1)}.\end{aligned}$$

Hence, for $c = \frac{1}{-\ln\mu_{2C}(1)} + \epsilon$, noting that $\mu_{2C}(1) < 1$ so $\ln\mu_{2C}(1) < 0$, we conclude

$$\mathsf{P}\left[\boldsymbol{\rho}(\ln\ln m) > c\ln m\right] = m^{-1+\epsilon\ln\mu_{2C}(1)+o(1)} + m^{-2+o(1)} = m^{-1-\Omega(1)}.$$

□

8.5.3 Phase 2: $\ln\ln m < \mathbf{r}_t < \chi$

Next we consider neighborhood sizes $\ln\ln m$ up to $\chi = m^{1/2}\ln^2 m$. In this range a "good" iteration occurs if $\mathbf{r}_{t+1} > M \cdot \mathbf{r}_t$ for arbitrary but fixed $M < M(\mu)$. The argument is similar to phase 1, except that in this case, if $\boldsymbol{\tau}(\ln\ln m) < \infty$, then the monotonicity of perimeter sizes implies that $\mathbf{r}_t > 0$ for the duration of the stopped BFS. Note also that the stopping time $\mathbf{T}$ will in fact be equal to $\boldsymbol{\tau}(\chi)$, except in cases where $\boldsymbol{\tau}(\chi) > \ln^2 m$.

Lemma 8.5.5. *For all $\epsilon > 0$, there exists $\delta > 0$ such that,*

$$\mathsf{P}\left[\infty > \boldsymbol{\tau}(\chi) > \boldsymbol{\tau}(\ln\ln m) + \left(\frac{1}{2M(\mu)} + \epsilon\right)\ln m\right] = m^{-1-\Omega(1)}$$

Proof. We begin by considering an arbitrary constant M satisfying $1 < M < M(\mu)$; the value of M will be specified further on. We now define

$$\mathbf{y}_t = \begin{cases} 1 & \text{if } \mathbf{r}_t/\mathbf{r}_{t-1} \geq M \text{ and } \mathbf{r}_{t-1} > \ln\ln m - 2 \\ 0 & \text{otherwise,} \end{cases}$$

and $\mathbf{Y}_t = \sum_{i=\boldsymbol{\tau}(\ln\ln m)}^{t} \mathbf{y}_i$. Corollary 8.5.3 implies that, with probability $1 - m^{-2+o(1)}$, $\mathbf{r}_t > \ln\ln m - 2$ for all $\boldsymbol{\tau}(\ln\ln m) < t < \mathbf{T}$, which in turn implies, again with probability $1 - m^{-2+o(1)}$, that

$$\mathbf{r}_t \geq (\ln\ln m - 2)M^{\mathbf{Y}_t} > e^{\mathbf{Y}_t \ln M}$$

assuming of course that m is sufficiently large so that $\ln\ln m - 2 > 1$.

Now, since $\boldsymbol{\rho}(\ln\ln m) = O(\ln m)$ with probability $1 - m^{-1-\Omega(1)}$ then $\boldsymbol{\tau}(\ln\ln m)$ is either $O(\ln m)$ or infinite, also with probability $1 - m^{-1-\Omega(1)}$. And, as noted above if $\boldsymbol{\tau}(\ln\ln m) = \infty$ then $\boldsymbol{\tau}(\chi) = \infty$ as well, in which case $\boldsymbol{\rho}(\ln\ln m) = \boldsymbol{\rho}(\chi)$. In either case, for any $s \leq \ln^2 m - O(\ln m)$ (since $\ln^2 m$ is the maximum duration before halting), we have

$$\mathsf{P}\left[\left(\boldsymbol{\rho}(\chi) > \boldsymbol{\tau}(\ln\ln m) + s\right) \wedge \left(\mathbf{Y}_{\boldsymbol{\rho}(\ln\ln m)+s} > \frac{\ln\chi}{\ln M}\right)\right] = 1 - m^{-1-\Omega(1)}. \tag{8.17}$$

Using the large deviations bound in statement 4 of lemma 8.5.2, we now deduce that

$$\mathsf{P}_t\left[\mathbf{y}_{t+1} = 0\right] \leq e^{-\Omega(\ln\ln m)} = o(1),$$

and the convergence implicit in the "$o(1)$" is uniform for all $t < \mathbf{T}$. Hence, by (8.17) and proposition 8.3.4, we compute

$$\mathsf{P}\left[\boldsymbol{\rho}(\chi) > \boldsymbol{\rho}(\ln\ln m) + s\right] \leq \mathsf{P}\left[\mathbf{x}_{\mathrm{Bin}_{(s,1-o(1))}} \leq \frac{\ln\chi}{\ln M}\right] + m^{-1-\Omega(1)}.$$

Since

$$\ln\chi = \ln m^{1/2}\ln^2 m = (1/2 + o(1))\ln m,$$

we set $s = c\ln m$ for a constant $c > \frac{1}{2\ln M}$, and deduce

$$\begin{aligned}
\mathsf{P}\left[\boldsymbol{\rho}(\chi) - \boldsymbol{\rho}(\ln\ln m) > c\ln m\right] &\leq \mathsf{P}\left[\mathbf{x}_{\mathrm{Bin}_{(c\ln m,1-o(1))}} \leq \left(\frac{1}{2\ln M} + o(1)\right)\ln m\right] + m^{-1-\Omega(1)} \\
&= \mathsf{P}\left[\mathbf{x}_{\mathrm{Bin}_{(c\ln m,o(1))}} \geq \left(c - \frac{1}{2\ln M} - o(1)\right)\ln m\right] + m^{-1-\Omega(1)} \\
&= \mathsf{P}\left[\mathbf{x}_{\mathrm{Bin}_{(c\ln m,o(1))}} \geq \Omega(\ln m)\right] + m^{-1-\Omega(1)} \\
&= m^{-1-\Omega(1)}.
\end{aligned}$$

Moreover, by making $M(\mu) - M$ arbitrarily small, we can achieve this bound for any $c = \frac{1}{2\ln M(\mu)} + \epsilon$.

To derive the claimed bound on $\boldsymbol{\tau}(\chi)$, note that if $\boldsymbol{\tau}(\ln\ln m) < \infty$, then $\mathbf{r}_{\boldsymbol{\tau}(\ln\ln m)} > \ln\ln m$, which implies that, with probability $1 - m^{-2+o(1)}$, $\mathbf{r}_t > 0$ for the duration of the stopped BFS process, which means that $\mathbf{r}_{\boldsymbol{\tau}(\chi)} > 0$, and hence it must be that case that $\boldsymbol{\tau}(\chi) = \boldsymbol{\rho}(\chi) < \infty$. □

8.5.4 The Case $\boldsymbol{\tau}(\chi) = \infty$

Lemma 8.5.6. *$\boldsymbol{\tau}(\chi) < \infty$ with probability at most $m^{-1+o(1)}$, and, for any fixed $\epsilon > 0$, with probability least $m^{-1-\Omega(1)}$, $\boldsymbol{\tau}(\chi) = \infty$ implies that $\boldsymbol{\rho} < \epsilon\ln m$.*

Proof. Recall that, with probability $1 - m^{-2+o(1)}$, $\mathbf{r}_t$ can only decrease once during the halted BFS, and the decrement can have size at most 2. It follows that if $\mathbf{r}_t$ ever reaches 3, then it must remain strictly positive for the duration. And, in this case the two lemmas for phases 1 and 2 imply that $\boldsymbol{\tau}(\chi) < \infty$.

Moreover since phase 1 cannot last more that $O(\ln m)$ iterations, then if $\mathbf{r}_t$ drops to 0 before $\boldsymbol{\tau}(\chi)$, this must happen in the first $O(\ln m)$ iterations. Since the probability of a drop from either $\mathbf{r}_t = 2$ or $\mathbf{r}_t = 1$ to $\mathbf{r}_{t+1} = 0$ is $O(1/m)$, then the probability that this occurs at all is at most $O(\ln m/m) = m^{-1+o(1)}$.

However, the fact that $\mathbf{r}_t$ must remain bounded by 2 until $\boldsymbol{\rho}$ imposes even more severe restrictions. Specifically, the sequence of perimeter sizes must begin with either $\mathbf{r}_0 = 1$ or $\mathbf{r}_0 = 2$, and there can only be one increase from $\mathbf{r}_t = 1$ to $\mathbf{r}_{t+1} = 2$.

For fixed $\epsilon > 0$, in order for $\boldsymbol{\rho} > \epsilon \ln m$ and $\boldsymbol{\tau}(\chi) = \infty$, again since $\boldsymbol{\rho} = O(\ln m)$ in this situation, then, based on the choice of $\boldsymbol{\rho}$ as well as the time of the single possible increase in $\mathbf{r}_t$, there are only $O(\ln^2 m)$ possible sequences of perimeter sizes $(\mathbf{r}_0, \ldots, \mathbf{r}_{\boldsymbol{\rho}})$.

However, every one of these sequences must include at least $\epsilon \ln m$ repeated perimeter sizes, which has probability $\psi'(z_\mu)^{\epsilon \ln m} = m^{-\Omega(1)}$, and also a drop to 0, which has probability $m^{-1+o(1)}$. Hence in total, all possibilities combined have probability at most $m^{-1-\Omega(1)+o(1)} \ln^2 m = m^{-1-\Omega(1)}$. □

The proof of theorem 8.5.1 is now essentially complete.

Proof of Theorem 8.5.1. Lemmas 8.5.4 and 8.5.5 jointly establish that, with probability $m^{-1-\Omega(1)}$, if $\boldsymbol{\tau}(\chi) < \infty$, then $\boldsymbol{\tau}(\chi) < \left(\frac{\Delta_\mu}{2} + \epsilon\right) \ln m$. Both claims of theorem 8.5.1 then follow directly from lemma 8.5.6. □

8.6 Upper Bound Proof of Theorem 8.2.1

The upper bound proof of Theorem 8.2.1 is easily pieced together from Theorem 8.5.1.

Theorem 8.6.1. *Assume $M(\mu) > 1$ and let $\mathbf{G} = \boldsymbol{\mathcal{G}}(\mu)$. Then, for any $\epsilon > 0$, the diameter of $\mathbf{G}$ is w.p.h.p. at most $(\Delta_\mu + \epsilon) \ln m$. More specifically, the following hold w.p.h.p.:*

1. *the maximum finite distance between any two vertices in the giant component is at most $(\Delta_\mu + \epsilon) \ln m$;*

2. *the maximum finite distance between any two vertices which belong to neither the giant component nor a tree component is at most* $\epsilon \ln m$

3. *the maximum distance between any two vertices in any tree component is at most* $(\frac{1}{-\ln \psi'(z_\mu)} + \epsilon) \ln m$.

8.6.1 The Giant Component and the 2-core

We first consider vertices which are connected to the 2-core, including those in the giant component.

Proof of Claims 1 and 2. Due to proposition 8.1.1 from section 8.1, the distance between any pair u, v such that both $\boldsymbol{\tau}(u, \chi) < \infty$ and $\boldsymbol{\tau}(v, \chi) < \infty$ is at most $\boldsymbol{\tau}(u, \chi) + \boldsymbol{\tau}(v, \chi) + 1$ w.p.h.p., and this holds simultaneously for all pairs.

As shown in proposition 8.1.3, 2-core BFS perimeters are contained in ordinary BFS perimeters, and therefore $\boldsymbol{\tau}(v, \chi) \leq \boldsymbol{\tau}_{2C}(v, \chi)$, so the hitting times $\boldsymbol{\tau}_{2C}(v, \chi)$ can also be used to achieve this upper bound.

Now, by proposition 8.1.5, the initial state of the 2-core BFS process for any vertex v w.e.h.p. satisfies assumptions 8.5.1, assuming that v is connected to the 2-core. Hence, for any fixed $\epsilon > 0$, by theorem 8.5.1 with probability $1 - m^{-1-\Omega(1)}$, any such vertex satisfies either $\boldsymbol{\rho}_{2C}(v) < \epsilon \ln m$ or $\boldsymbol{\tau}_{2C}(v, \chi) < \left(\frac{\Delta_\mu + \epsilon}{2}\right) \ln m$.

Since this bound in probability of $1 - m^{-1-\Omega(1)}$, holds uniformly over all vertices, then it holds simultaneously for all vertices with probability $1 - m^{-\Omega(1)}$, which is polynomially high. In this case, we have an immediate bound on the distance between any pair of vertices both of which satisfy $\boldsymbol{\tau}_{2C}(v, \chi) < \infty$.

It is not hard to see that any vertex with $\boldsymbol{\tau}_{2C}(v, \chi) < \infty$ belongs to the giant component; indeed, any pair of such vertices satisfies $\boldsymbol{\delta}(u, v) < \infty$, and by theorem 8.5.1, all but $m^{-1+o(1)}$ of the vertices which are connected to the 2-core also satisfy this property. Since the 2-core itself has size $\Omega(m)$, it follows that there are $\Omega(m)$ such vertices, and hence this must be the giant component.

Conversely, if $\boldsymbol{\tau}_{2C}(v,\chi) = \infty$, then $\boldsymbol{\rho}_{2C}(v) < \epsilon \ln m$, which means that v is only connected by a path to at most $\epsilon \ln m$ vertices in the 2-core. Such a vertex does not belong to the giant component, and so if $\boldsymbol{\tau}_{2C}(u,\chi) < \infty$, then u and v belong to different components, so we have $\delta(u,v) = \infty$.

In the third case, if both $\boldsymbol{\tau}_{2C}(u,\chi) = \boldsymbol{\tau}_{2C}(v,\chi) = \infty$, it may be the case that these vertices are connected by a path. But, assuming that both of these are connected to the 2-core, the bound $\boldsymbol{\rho}_{2C}(v) < \epsilon \ln m$, implies that the most distant 2-core vertex from v is at distance at most $\epsilon \ln m$, and therefore $\boldsymbol{\delta}(u,v) < 2\epsilon \ln m$ in this case. □

8.6.2 Tree Components

We now bound the diameter of tree components.

Proof of Claim 3. Note that if u, v belong to the same tree component, then then the augmented 2-core endpoint set $\mathbf{A}_{2C,\{u,v\}}$ consists of the ordinary 2-core endpoint set $\mathbf{A}_{2C}$, along with a path connecting u, v which is disjoint from $\mathbf{A}_{2C}$. Also, the unexposed degrees of u, v must satisfy

$$\mathbf{deg}_{2C,\{u,v\}}(u) = \mathbf{deg}_{2C,\{u,v\}}(v) = 1.$$

Now, if we perform BFS beginning from u in the augmented 2-core $\mathbf{G}_{2\mathcal{C},\{u,v\}}$, in order to reveal a path of length r connecting u to v which does not connect to the 2-core, we must encounter $r-1$ consecutive endpoints of unexposed residual degree 1, and then in the r'th iteration, we must encounter the remaining unexposed endpoint on the vertex v. Since the fraction of endpoints of unexposed residual degree 1 is at w.e.h.p. at most $\mu_{2C}(1)+o(1)$, then the probability of choosing $\left(\frac{1}{-\ln \mu_{2C}(1)} + \epsilon\right) \ln n$ consecutive endpoints of residual degree 1 is $m^{-1-\Omega(1)}$. And, the probability of then choosing the remaining endpoint on v is $O(m^{-1})$, so the overall probability is $m^{-2-\Omega(1)}$. The lemma now follows by the considering all $O(n^2)$ possible pairs of endpoints and invoking the first moment method. □

8.7 Lower Bound Proof of Theorem 8.2.1

In this section we prove the lower bound of Theorem 8.2.1, which, in light of the 2-core path decomposition, is somewhat easier than the upper bound proof.

Theorem 8.7.1. *Assume $M(\mu) > 1$ and let $\boldsymbol{\mathcal{G}}(A,V) = \boldsymbol{\mathcal{G}}(\mu)$, and assume also that $M(\mu_{(A,V)}) \to M(\mu)$. Then, for any $\epsilon > 0$, the diameter of $\mathbf{G}$ is w.p.h.p. at least $(\Delta_\mu - \epsilon)\ln m$.*

8.7.1 Upper Bound on Neighborhood Sizes

The first step in the lower bound proof is to derive an upper bound on the rate of neighborhood growth in a BFS process. The upper bound is different in nature than the lower bound derived in section 8.4 in two ways. First, we only use the expected neighborhood size to compute the bound. Second, we bound the probability that the sum of all neighborhoods up to time t exceeds a given amount, rather than considering individual iterations.

Lemma 8.7.2. *For any $1 < M < \infty$, and any $\epsilon > 0$, there exists a $\delta > 0$ such that the following statement holds.*

For any endpoint partition (A,V) with $|A| = m$ and such that $M(\mu_{(A,V)}) < M + \delta$, and for any $q < \delta m$ and $t \geq 0$, we have

$$\mathsf{P}\left[\sum_{i=0}^{t} \mathbf{r}_t(v) > q\right] < \frac{\deg(v)(M+\epsilon)^t}{q(1-1/M)}.$$

Proof. For any $i \geq 0$, note that $\mathbf{r}_{t+1}(v)$ is at most equal to the sum of the residual degrees of the endpoints $\vec{\mathbf{E}}(a)$ for $a \in \mathbf{R}_t(v)$ (with equality if no cross-edges occur). Hence, by linearity of expectation, we have

$$\mathsf{E}_t[\mathbf{r}_{t+1}(v)] \leq \sum_{a \in \mathbf{R}_t(v)} \mathsf{E}_t\left[\mathrm{res}(\vec{\mathbf{E}}(a))\right]$$

Now for any given a, we compute

$$\mathsf{E}_t\left[\mathrm{res}(\vec{\mathbf{E}}(a))\right] = \frac{\sum_{a_0 \in \mathbf{A}_t(v)\setminus\{a\}} \mathrm{res}(a_0)}{|\mathbf{A}_t(v)| - 1} \leq \frac{\sum_{a_0 \in A} \mathrm{res}(a_0)}{|\mathbf{A}_t(v)| - 1} = \left(\frac{m}{|\mathbf{A}_t(v)| - 1}\right) M(\mu_{(A,V)}),$$

and therefore, for δ sufficiently small, we can guarantee $\mathsf{E}_t[\mathbf{r}_t(v)] \le \mathbf{r}_t(v)(M+\epsilon)$ whenever both $M\left(\mu_{(A,V)}\right) < M+\delta$ and $\sum_{j=0}^{i} \mathbf{r}_j(v) < \delta m$.

We now define the random variables

$$\mathbf{q}_t(v) = \begin{cases} \mathbf{r}_t(v) & \text{if } \sum_{j=0}^{t-1} \mathbf{r}_{i-1}(v) < \delta m, \\ 0 & \text{otherwise,} \end{cases}$$

and we deduce that $\mathsf{E}_t\left[\mathbf{q}_{t+1}(v)\right] \le \mathbf{q}_t(v)(M+\epsilon)$ always, and therefore $\mathsf{E}\left[\mathbf{q}_t(v)\right] \le \deg(v)(M+\epsilon)^t$ for all i.

For $q < \delta m$, using Markov's inequality, we now compute

$$\mathsf{P}\left[\sum_{i=0}^{t} \mathbf{r}_t(v) > q\right] = \mathsf{P}\left[\sum_{i=0}^{t} \mathbf{q}_t(v) > q\right] \le \frac{1}{q}\mathsf{E}\left[\sum_{i=0}^{t} \mathbf{q}_t(v)\right] \le \frac{\deg(v)}{q}\sum_{i=0}^{t}(M+\epsilon)^i \le \frac{\deg(v)(M+\epsilon)^t}{q(1-1/M)}.$$

□

8.7.2 Average Distance in the 2-core

Next, we first prove that almost all pairs of vertices in a graph with residual distribution μ_{2C} are connected by a path of length at least $\frac{\ln n}{\ln M(\mu)} - o(\ln n)$.

Lemma 8.7.3. *Let $\boldsymbol{\mathcal{G}}(A,V) = \boldsymbol{\mathcal{G}}(\mu_{2C})$, and assume that $M(\mu_{(A,V)}) \to M(\mu) = M(\mu_{2C})$, and also that the minimum positive degree is at least 2.*

For any pair of vertices u, v of degree $O(1)$, and for any fixed $\epsilon > 0$

$$\mathsf{P}\left[\infty > \boldsymbol{\delta}(u,v) > \left(\frac{1}{\ln M(\mu)} - \epsilon\right)\ln m\right] = 1 - m^{-\Omega(1)}.$$

Proof. The upper bound proof of the diameter shows that the probability that any given vertex does not belong to the giant component is $m^{-1+o(1)}$. Hence, for any given pair $u, v \in V$, we have $\mathsf{P}[\boldsymbol{\delta}(u,v) < \infty] = 1 - m^{-1+o(1)}$.

For the lower bound on $\boldsymbol{\delta}(u,v)$, we use the upper bound on neighborhood growth from lemma 8.7.2. Since $M(\mu_{2C}) = M(\mu)$, then for arbitrary $\epsilon_0 > 0$ there exists $\delta > 0$ such that

$$\mathsf{P}\left[\sum_{i=0}^{t} \mathbf{r}_t(v) > q\right] = O\left(\frac{(M(\mu)+\epsilon_0)^t}{q}\right)$$

for all $q < \delta m$.

By setting $q = m^c$ for an appropriate choice of $0 < c < 1$ and by making ϵ_0 sufficiently small, for all $t \leq \left(\frac{1}{\ln M(\mu)} - \epsilon\right) \ln n$ we may deduce that $\mathsf{P}\left[\sum_{i=0}^{t} \mathbf{r}_t(v) > m^{1-\Omega(1)}\right] = 1 - m^{-\Omega(1)}$.

In other words, with probability $1 - m^{-\Omega(1)}$, the first t neighborhoods have combined size $m^{1-\Omega(1)}$, and thus the probability that a path from v to an arbitrary vertex of degree $O(1)$ is exposed is also $m^{-\Omega(1)}$. We conclude that the distance from u to v is at least t with probability $1 - m^{-\Omega(1)}$. □

8.7.3 Separation From the 2-core

Our final lemma shows that we can find vertices which are separated from the 2-core by a large distance.

Lemma 8.7.4. *For any $0 < c < \frac{1}{-\ln \psi'(z_\mu)}$, w.e.h.p. there exist at least two vertices u, v such that*

1. *both u, v are connected to the 2-core by a path of length at least $c \ln m$;*
2. *the closest 2-core vertex to v is different from the closest 2-core vertex to u, and in both cases the degree of this closest vertex is $O(1)$.*

Proof. We examine the behavior of the 2-core process in detail near the termination time. Hence, let us stop the 2-core algorithm at a first time $\boldsymbol{\tau}_0$ such that the number of endpoints of degree 1 drops below $m^{1-\delta}$, where $\delta > 0$ is an arbitrary constant we shall choose further on.

Since the number of endpoints is of residual degree 1 is $o(m)$, using the same argument as for the augmented 2-core in proposition 8.1.5, we deduce that the residual distribution at this time satisfies $\left\|\boldsymbol{\mu}_{\boldsymbol{\tau}_0} - \mu_{2C}\right\|_1 = o(1)$, w.e.h.p.

We now customize the 2-core process as follows. We choose a vertex v of degree 1, and we expose the entire path from v to the closest vertex of any degree other than 2. Similar

to the argument in the upper bound proof, the fact that $\boldsymbol{\mu}_{\boldsymbol{\tau}_0}(1) = \psi'(z_\mu) \pm o(1)$ implies that the length of the path exposed in this way is at most $O(\ln m)$ with probability $m^{-1-\Omega(1)}$. Also, after this path has been found, the number of vertices of degree 1 drops by either 1 or 2, depending on whether the terminal vertex has degree 1; vertices along this path change from degree 2 to degree 0, so in the end no new vertices of degree 1 are created.

It follows that if we perform this operation repeatedly, by successively choosing initial vertices $(\mathbf{v}_1, \mathbf{v}_2, \ldots)$ of degree 1, we have at least $m^{1-\delta}/2$ different initial vertices, and the total number of endpoints exposed is w.p.h.p. $m^{1-\delta} \ln m = m^{1-\Omega(1)}$. Let us now define an indicator random variable $\mathbf{z}_t$ to count the number of times that

1. the path revealed starting with $\mathbf{v}_t$ has length at least $c \ln m$;
2. the terminal vertex has degree in the range $\{3, \ldots, J\}$, for some constant J.

Note that the probability of exposing a path of length $c \ln m$ or greater is at least

$$(\psi'(z_\mu) - o(1))^{c \ln m} = m^{c \ln \psi'(\mu) - o(1)} = m^{-c_1},$$

for some constant $\frac{c}{-\ln \psi'(z_\mu)} < c_1 < 1$. Moreover, by choosing J appropriately, we can ensure that a constant fraction of the remaining endpoints will belong to vertices of degrees between 3 and J, it follows that c_1 can be chosen such that

$$\mathsf{P}_t[\mathbf{z}_{t+1} = 1] > m^{-c_1}$$

for the duration of this process

It follows that the total number of occurrences of $\mathbf{z}_t$ dominates the binomial distribution $\mathrm{Bin}_{(m^{1-\delta}/2, m^{-c_1})}$, and, since $c_1 < 1$, then by choosing $\delta > 0$ sufficiently small, we may conclude that w.p.h.p. there are at least two such occurrences.

Finally, w.l.o.g. assume that $\mathbf{v}_1, \mathbf{v}_2$ are the first two vertices for which the above event occurs. Note that the probability that both end up connected to the same terminal vertex is $O(1/m)$, since this vertex has constant degree. Moreover, since only $m^{1-\Omega(1)}$ additional endpoints are exposed until this process terminates, then the probability that no additional

endpoints on either of these terminal vertices becomes exposed for the duration of this process is $1 - m^{-\Omega(1)}$, and hence w.p.h.p. two such vertices exist. □

Proof of Theorem 8.7.1. By the above lemma, for arbitrary $\epsilon > 0$, w.p.h.p. there exist two distinct vertices u, v which belong to the 2-core, have degree $O(1)$ in the 2-core, and are connected by a path of length at least $\left(\frac{1}{-\ln \psi'(z_\mu)} - \epsilon\right) \ln m$ to a vertex of degree 1 outside the 2-core (and such that this path does not belong to the 2-core).

Moreover, since the 2-core edge set is uniform conditional on the 2-core endpoint set, then the fact that these vertices satisfy these properties does not affect the distribution of the random variable $\boldsymbol{\delta}(u, v)$, conditional on $\mathbf{A}_{2C}$. In this case, by lemma 8.7.3, $\boldsymbol{\delta}(u, v)$ is w.p.h.p. finite and at least $\left(\frac{1}{M(\mu)} - \epsilon\right) \ln m$ and it follows that the two corresponding vertices of degree 1 are at distance $(\Delta_\mu - \epsilon) \ln m$. □

8.8 Other Cases

We now consider generalizations of Theorem 8.2.1 to graphs of higher minimum degree. The proofs of the generalizations in this section are quite similar to the proof of Theorem 8.2.1 given above. Hence we only offer proof sketches which describe how to adapt the original proof to handle particular generalizations.

Recall that the term $\frac{2}{-\ln \mu_{2C}(1)}$ in the constant Δ_μ reflects the fact that the longest shortest path will include two long "strands" of length $\frac{1}{-\ln \mu_{2C}(1)}$. Moreover, the length of these strands is determined by the fact that the probability of repeating a neighborhood size during 2-core BFS is $\mu_{2C}(1) \pm o(1)$ for a neighborhood of size 1.

The length of these "strands" changes when the minimum degree is at least 2. First, obviously, the 2-core of a graph with minimum degree at least 2 is simply the entire graph, so $\mu_{2C}(1) = \mu(1)$. More significantly, for graphs with minimum degree exactly 2, the initial neighborhood size is also at least 2, so in order to repeat a neighborhood size, in general we must choose 2 consecutive endpoints of residual degree 1. As a result, the longest "strands" will have length $\frac{1}{-2 \ln \mu(1)}$. For graphs with minimum degree at least 3, there are

no endpoints of residual degree either 0 or 1, and hence a neighborhood size can only be repeated when cross-edges occur. In this case, the "strands" have length $o(\ln n)$, and the diameter is determined by $\frac{1}{\ln M(\mu)}$. We now consider these two cases formally in the two theorems stated below.

First, we consider graphs with $\mu(0) = 0$ and $\mu(1) > 0$. In this case, our asymptotic assumptions are not sufficient to determine the diameter; we must also require that the minimum degree is in fact 2, otherwise the presence of vertices of degree 1 may increase the diameter.

Theorem 8.8.1. *Let $\boldsymbol{\mathcal{G}}(A, V) = \boldsymbol{\mathcal{G}}(\mu)$ where $0 < \mu(1) < 1$ and $\mu(0) = 0$, and assume that $M(\mu_{(A,V)}) \to M(\mu)$ and and that the minimum degree in (A, V) is 2. Then w.p.h.p.*

$$\frac{\Delta(\boldsymbol{\mathcal{G}}(A, V))}{\ln m} \to \frac{1}{-\ln \mu(1)} + \frac{1}{\ln M(\mu)}.$$

Proof Sketch. For the upper bound, we note that the first endpoint neighborhood during BFS contains at least 2 endpoints, and the probability of repeating a neighborhood size of 2 is $\mu(1)^2 \pm o(1)$. Using this observation, it is straightforward to adapt the proof of lemma 8.5.4 to show that for any v and any $\epsilon_0 > 0$, we have

$$\boldsymbol{\rho}(\ln \ln m) < \left(\frac{1}{-2 \ln \mu(1)} + \epsilon_0\right) \ln m$$

with probability $1 - m^{-1-\Omega(1)}$. The only minor technicality is the fact that it is possible to encounter a situation where the neighborhood size drops to 1 due to cross-edges. However, the probability of such an occurrence is $O(m^{-1})$ for any given BFS iteration, and hence the probability that an endpoint neighborhood of size 1 is either preceded or followed by more than $\epsilon_0 \ln m$ non-empty neighborhoods of size less than $\ln \ln m$ is $m^{-1-\Omega(1)}$.

The lower bound proof can be modified similarly by finding a pair of vertices u, v, such that both u and v have $\left(\frac{1}{-2 \ln \mu(1)} - \epsilon_0\right) \ln m$ consecutive endpoint neighborhoods of size 2. □

We now consider graphs with $\mu(0) = \mu(1) = 0$ and with minimum degree 3.

Theorem 8.8.2. *Let $\boldsymbol{\mathcal{G}}(A,V) = \boldsymbol{\mathcal{G}}(\mu)$ where $\mu(0) = \mu(1) = 0$, and assume that $M(\mu_{(A,V)}) \to M(\mu)$ and and that the minimum degree in (A,V) is 3. Then w.p.h.p.*

$$\frac{\Delta(\boldsymbol{\mathcal{G}}(A,V))}{\ln m} \to \frac{1}{\ln M(\mu)}.$$

Proof Sketch. In this case, the minimum degree is at least 3, and hence the minimum residual degree is at least 2. It follows that, for a neighborhood of size $\mathbf{r}_t(v) = O(\ln\ln n)$, we have $\mathbf{r}_{t+1}(v) > \mathbf{r}_t(v)$ with probability $1 - o(1)$ and from this we can deduce $\boldsymbol{\rho}(\ln\ln n) = o(\ln n)$ with probability $1 - m^{-1-\Omega(1)}$, and the upper bound follows. And, in this situation the lower bound follows directly from lemma 8.7.3. □

8.9 Computing the Diameter for Specific Degree Distributions

In this section, we compute the specific value of

$$\Delta_\mu = \frac{2}{-\ln \psi'_\mu(z_\mu)} + \frac{1}{\ln M(\mu)}$$

from theorem 8.2.1 for two specific classes of distributions. The first is the Poisson distribution, which corresponds to the Erdős-Rényi random graphs $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$, and the second is the power-law distribution which is used in that power-law graph model of Aiello, Chung, and Lu [2].

8.9.1 The Diameter of $\mathbf{G}_{n,p}$

We now consider the diameter of the classical random graph $\mathbf{G}_{n,p}$, for $p = \frac{d}{n}$, where $d > 1$. As discussed in chapter 3, and in the appendix A, $\mathbf{G}_{n,p}$ can be simulated by specifying a Poisson degree distribution π_d, in which case the residual distribution is also Poisson, and hence the diameter of $\mathbf{G}_{n,p}$ is given w.p.h.p. by

$$\Delta(\mathbf{G}_{n,d/n}) = \Delta_{\pi_d} \ln n \pm o(\ln n).$$

The p.g.f. for the Poisson distribution π_d has the simple expression $\psi_{\pi_d}(z) = e^{d(z-1)}$. The fixed point of this function is given by $z_{\pi_d} = \frac{-W(-de^{-d})}{d}$, where the *Lambert W-function*

$W(z)$ is the principal inverse of $f(z) = ze^z$. This gives the following expression for Δ_{π_d}:

$$\Delta_{\pi_d} = \frac{2}{\ln -W(de^{-d})} + \frac{1}{\ln d}. \tag{8.18}$$

The "average distance" between vertices in the giant component of $\mathbf{G}_{n,d/n}$ is simply $\ln n / \ln d = \log_d n$, since the average residual degree is $M(\pi_d) = d$. The actual diameter hence exceeds the average distance by $\frac{2 \ln n}{\ln -W(de^{-d})}$. The qualitative behavior of the diameter of sparse $\mathbf{G}_{n,p}$ can be understood by examining the plot in figure 8.1, which shows the ratio of the diameter to the average distance as a function of the average degree.

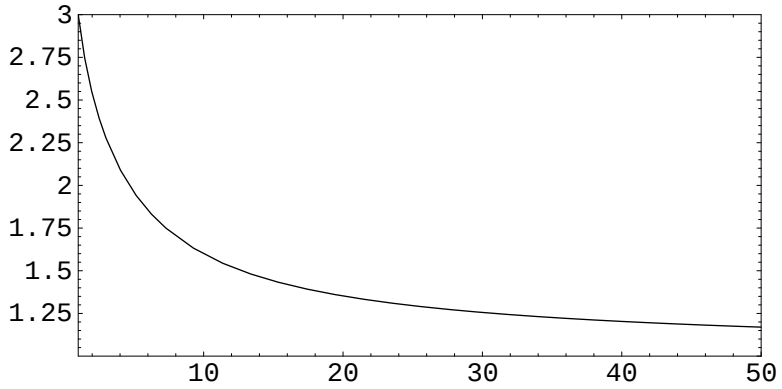


Figure 8.1: The function $\frac{\Delta_{\pi_d}}{1/\ln d}$ as a function of the average degree d, which measures the ratio of the diameter to the average distance in the random graph $\mathbf{G}_{n,d/n}$.

From equation (8.18), it can be shown that $\Delta_{\pi_d} \ln d \to 3$ as $d \to 1$ and $\Delta_{\pi_d} \ln d \to 1$ as $d \to \infty$ as the above plot suggests, and it is a simple exercise to derive increasingly accurate asymptotic characterizations of Δ_{π_d}, as in

$$\Delta_{\pi_d} = \frac{1}{\ln d} + \frac{2}{d} + O\left(\frac{\ln d}{d^2}\right) \text{ as } d \to \infty$$

and so on.

8.9.2 The Diameter of Power Law Graphs

We now compute the diameter of random graphs with so-called "power-law" distributions, in which the number of vertices of degree d is proportional to $d^{-\beta}$ for a given constant

β. The degree distribution for the random power law graph model proposed by [2] is

$$\lambda_\beta(i) = \frac{i^{-\beta}}{\zeta(\beta)},$$

with the zeta function $\zeta(\beta) = \sum_{n=1}^{\infty} n^{-\beta}$. The residual distribution is therefore given by

$$\mu_\beta(i) = \frac{(i+1)^{1-\beta}}{\zeta(1-\beta)},$$

and the corresponding p.g.f. is

$$\psi_{\mu_\beta}(z) = \frac{\mathrm{Li}(\beta-1, z)}{z\zeta(\beta-1)},$$

with the polylogarithm $\mathrm{Li}(x, z) = \sum_{n=1}^{\infty} \frac{z^n}{n^x}$.

Note that both the average degree and the average residual degree decrease as β increases. For $\beta > 2$, the average degree is $O(1)$, so the graph is sparse, and for $\beta > 3$, the average residual degree is $O(1)$ so the "average distance" is $\Omega(\ln n)$ In addition, the giant component phase transition for this model occurs at the point $\beta = 3.479...$, where $M(\mu_\beta) = 1$, so for $\beta > 3.479...$, the graph no longer exhibits a giant component.

Hence, the results of this chapter are relevant for the range $2 < \beta < 3.479$. Moreover, for $2 < \beta \leq 3$, the "average distance" is $o(\ln n)$, and therefore the leading constant for the diameter is simply given by $\frac{2}{-\ln \psi'_\mu(z_\mu)}$. The leading constant for the diameter of a random power law graphs is plotted in figure as a function of β, along with the function $\frac{2}{-\ln \psi'_\mu(z_\mu)}$ for $\beta > 3$.

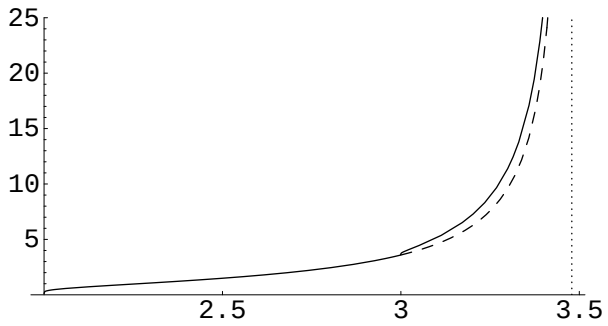


Figure 8.2: The leading constant Δ_μ for the diameter of random graphs generated by the power law model introduced in [2], as a function of β. For $\beta > 3$, the dashed line shows the contribution of long isolated paths, given by $\frac{2}{-\ln \psi'_\mu(z_\mu)}$. The dotted vertical line shows the phase transition, which occurs at $\beta = 3.479....$

Chapter 9

The k-Core

The k-core of the graph is the maximal induced sub-graph with minimum positive degree at least k, and in this chapter, we study the k-core of a random graph $\boldsymbol{\mathcal{G}}(A, V)$. Our method for analyzing the k-core is similar to the method employed in chapter 6 for determining the size and degree distribution of the 2-core: we trace the execution of a simple greedy k-core finding algorithm. As such, the analysis of the k-core returns us to the methodology developed in chapter 4 involving topological representation and differential equations, which was temporarily put on hold during our discussion of local structure and the diameter in the last two chapters.

Chapter Organization

The generalization from the 2-core to the k-core, which is presented in section 9.1, is fairly straightforward. Hence, in addition to just solving the k-core problem, we shall also develop some more general methods which will be used in the next chapter for the k-orientability problem.

In section 9.2, we develop certain variations of the probability generating function (p.g.f.) which allow us to state the solution to the k-core problem in an algebraic manner similar to the solution to the 2-core. Moreover, as with the 2-core, the solution to the k-core for $\boldsymbol{\mathcal{G}}(\mu)$ has a natural interpretation with respect to the corresponding Galton-Watson tree, and the variations of the p.g.f. can be used to describe both the graph property and the GW tree property succinctly.

Finally, in section 9.3, we discuss the possibility of finding a k-core of size $o(m)$. Unlike the 2-core, we are able to demonstrate, that, in most cases, the k-core for $k \geq 3$ must be either giant (of size $\Omega(m)$), or empty.

History and Background

The k-core problem for the Erdős-Rényi random graphs $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$ was first solved by Pittel, Spencer, in Wormald [59]. The k-core problem for general degree distributions was solved independently by Molloy [54], Cooper [22], and Janson and Luczak [38], and Fernholz and Ramachandran [32], all using similar techniques. The presentation in this chapter, and particularly the use of variants of the p.g.f. first appeared in [31], as part of the analysis of the related k-orientability algorithm which will be discussed in the next chapter.

9.1 The k-Core Process

The basic greedy algorithm for finding the k-core of a graph is almost identical to the 2-core algorithm described in chapter 6: we choose any vertex of degree $i < k$ in a graph G, remove this vertex along with all incident edges, and then recursively find the k-core in the residual graph.

The CM process corresponding to this algorithm will find all of the cores in a single process, which can be accomplished by removing an endpoint from a vertex of minimum degree each odd time step. This process is thus an extension of the 2-core process, since at first, all deterministically chosen endpoints will have unexposed degree 1. After the 2-core is found, the deterministic selections may have degree either 1 or 2, until the minimum degree reaches (at least) 3 at an even time, at which point the 3-core is found and so on.

We shall call such an endpoint removal process a *CM k-core process*, despite the fact that this process is not terminated, and finds all of the cores of a random graph rather than just one.

Algorithm 9.1.1 (CM k-core Process)**.** *Given an (even) endpoint partition (A, V), let $\mathbf{A}_0 = A$, and at each time step, remove an endpoint $\mathbf{a}_{t+1}$ from $\mathbf{A}_t$ to yield $\mathbf{A}_{t+1} = \mathbf{A}_t \backslash \{\mathbf{a}_{t+1}\}$ as follows:*

- *if t is* even*, choose an endpoint $\mathbf{a}_{t+1} \in \mathbf{A}_t$ with minimum unexposed degree;*

- *if t is* odd, *choose $\mathbf{a}_{t+1}$ uniformly at random from $\mathbf{A}_t$;*

In analyzing this process, we let $\mathbf{dmin}_t$ denote the minimum (positive) unexposed degree at time t, and hence we have $\mathbf{deg}_t(\mathbf{a}_{t+1}) = \mathbf{dmin}_t$ for all even values of $t < m$. In addition to the above definitions, we define random times

$$\boldsymbol{\tau}[k] = \min\{t : t \text{ is even and } \mathbf{dmin}_t \geq k\}$$

for each k; hence the k-core endpoint set is the set $\mathbf{A}_{\boldsymbol{\tau}[k]}$. And, as for the 2-core algorithm, conditional on $\mathbf{A}_{\boldsymbol{\tau}[k]}$ the k-core edge set is a uniformly distributed random matching of $\mathbf{A}_{\boldsymbol{\tau}[k]}$.

As for the 2-core, the k-core process makes use of two selection methods, D (deterministic) and R (random), which we describe in terms of indicator and measure processes ($\mathbf{I}_t[D]$) and $\Delta\boldsymbol{\sigma}_{t+1}[D] = \mathbf{I}_t[D]$. The state at time t determines the method by which $\mathbf{a}_{t+1}$ is selected, so the state is random (i.e. $\mathbf{I}_t[R] = 1$) if t is *odd*, and the state is deterministic (i.e. $\mathbf{I}_t[D] = 1$) if t is *even*.

9.1.1 Solution to the CM k-Core Process

The analysis of the k-core process is quite similar to the analysis of the 2-core process. Again we will work with the total random and deterministic weight rather the actual number of random and deterministic steps. Recall that the incremental weight of a step at time t is defined by $\Delta\mathbf{w}_t = \frac{1}{|\mathbf{A}_t|} = \frac{1}{m-t-1}$, and the solution to this process satisfies $e^{-w_\xi} = 1 - \xi$. We shall change variables as for the 2-core process by defining

$$\mathbf{x}_t = 1 - e^{-\mathbf{w}_t[R]}$$

$$\mathbf{y}_t = 1 - e^{-\mathbf{w}_t[D]}.$$

Since random and deterministic steps alternate, we again have

$$x_\xi = y_\xi = 1 - e^{-w_\xi/2} = 1 - \sqrt{1-\xi}.$$

Accordingly, given any initial residual distribution μ, we shall define values

$$\nu_{(x,y)}(i) = \left(\frac{1}{1-y}\right) \sum_{j=i}^{\infty} \mu(j)\, \mathrm{Bin}_{(j,1-x)}(i). \tag{9.1}$$

for $i \in \mathbb{Z}^*$, recalling that, by corollary 5.3.3 from chapter 5, for any endpoint removal process for which all deterministic removals all have degree at most k (or, more precisely, such that $d\sigma_\xi[D(i)] = 0$ for $i > k$), any w.e.h.p. solution to the residual distribution must satisfy

$$\mu_\xi(i) = \nu_{(x_\xi, y_\xi)}(i)$$

for all $i \geq k$. And, since random and deterministic steps alternate in the core process, we may simply define

$$\nu_\xi(i) = \nu_{(\sqrt{1-\xi}, \sqrt{1-\xi})}(i). \tag{9.2}$$

Now, all deterministic removals have degree at most $k-1$ before the time $\boldsymbol{\tau}[k]$. As usual, we scale this random time by dividing by the total duration $m = |A|$, and hence a solution will satisfy $\boldsymbol{\tau}[k]/m \to \tau[k]$. It follows that any w.e.h.p. solution must satisfy $\mu_\xi(i) = \nu_\xi(i)$ for all $i > k-1$ and all $\xi < \tau[k]$. Also, if $\tau[k] < 1$, this equation extends by continuity to $\xi = \tau[k]$, and thus we have the following proposition.

Proposition 9.1.1. *For any w.e.h.p. solution to the core process such that $\tau[k] < 1$, we have*

$$\mu_{\tau[k]}(i) = \nu_{\tau[k]}(i)$$

for $i \geq k-1$ and $\mu_{\tau[k]}(i) = 0$ for $i < k-1$.

Proof. By continuity, any solution with $\tau[k] < 1$ must also satisfy

$$\mu_{\tau[k]}(i) = \lim_{\xi \to \tau[k]^-} \mu_\xi(i) = \nu_\xi(i)$$

for all $i \geq k-1$. Also, since $\mathbf{dmin}_{\boldsymbol{\tau}[k]} = k$, then any solution to $\tau[k]$ must satisfy $\lambda_{\tau[k]}(i) = 0$ for $i < k$, and therefore, if $\tau[k] < 1$, we must have

$$\mu_{\tau[k]}(i) = \frac{(i+1)\lambda_{\tau[k]}(i+1)}{M(\lambda_{\tau[k]})} = 0$$

for $i < k-1$. □

This proposition yields the exact residual distribution $\mu_{\tau[k]}$ of the k-core, given the value of $\tau[k]$, and assuming that $\tau[k] < 1$. Moreover, the values $\mu_{\tau[k]}(i)$ must sum to 1, and it follows that

$$\sum_{i \in \mathbb{Z}^*} \mu_{\tau[k]}(i) = \sum_{i \geq k-1} \nu_{\tau[k]}(i) = 1.$$

Hence, all possible solutions to $\tau[k]$ must satisfy the above equation. In some cases, we may use this observation to find a unique solution to $\tau[k]$ as follows.

Proposition 9.1.2. *For any distribution μ, let us define, in terms of the values ν_ξ computed in (9.2), a function*

$$f_k(\xi) = \sum_{i \geq k-1} \nu_\xi(i) \tag{9.3}$$

and

$$\xi_k = \sup\{\xi \in [0,1] : f_k(\xi) > 1\}. \tag{9.4}$$

Then:

1. *every w.e.h.p. solution to the k-core process with initial residual distribution μ must satisfy $f_k(\tau[k]) = 1$, and hence $\tau[k] \geq \xi_k$;*
2. *if, in addition, the value ξ_k satisfies*

$$\xi_k = \inf\{\xi \in [0,1] : f_k(\xi) > 1\},$$

then $\tau[k] = \xi_k$ is the unique w.e.h.p. solution.

Proof. It is shown above that every solution to $\tau[k]$ must satisfy $f_k(\tau[k]) = 1$. Note also that $\nu_0(i) = \mu(i)$, and hence $f_k(0) = \sum_{i \geq k-1} \mu(i) \leq 1$. It follows that ξ_k is the smallest value such that $f_k(\xi) = 1$, and hence ξ_k is the smallest possible solution.

For the second claim, for any solution $\tau[k]$ and any $\xi < \tau[k]$, we have

$$\sum_{i \in \mathbb{Z}^*} \mu_\xi(i) \geq \sum_{i \geq k-1} \nu_\xi(i) = f_k(\xi).$$

In particular, the $\mu_\xi(i)$ cannot sum to a value greater than 1, hence if $f_k(\xi) > 1$ we must have $\tau[k] \leq \xi$, and it follows that $\tau[k] \leq \inf\{\xi \in [0,1] : f_k(\xi) > 1\}$. □

We may thus deduce the following about the k-core of a random graph $\boldsymbol{\mathcal{G}}(\mu)$.

Theorem 9.1.3. *Let $\boldsymbol{\mathcal{G}}(\mu)$ be a random graph with limiting residual distribution μ, and define f_k and ν_ξ as in (9.2) and (9.3). If there exists a value*

$$\xi_k = \sup\{\xi \in [0,1] : f_k(\xi) < 1\} = \inf\{\xi \in [0,1] : f_k(\xi) > 1\} \tag{9.5}$$

then:

1. *the k-core of $\boldsymbol{\mathcal{G}}(\mu)$ w.e.h.p. contains $(1-\xi_k)(m \pm o(m))$ endpoints;*
2. *the unique w.e.h.p. solution to the residual distribution of the k-core is given by $\mu_{\xi_k}(i) = \nu_{\xi_k}(i)$ for $i \geq k-1$ and $\mu_{\xi_k}(i) = 0$ for $i < k-1$.*

Proof. Immediate. □

This theorem allows us to compute the size and residual distribution of the k-core of a random graph $\boldsymbol{\mathcal{G}}(\mu)$, provided that the value ξ_k in (9.5) exists. The case where the supremum and infimum in this equation do not coincide corresponds to a threshold condition. In this situation, there are multiple w.e.h.p. solutions to $\tau[k]$, and hence the size of the k-core cannot be determined w.e.h.p. Moreover, it is fairly easy to see that this threshold is sharp, in the sense that the closure of the set of distributions for which ξ_k exists is the entire space Φ of distributions on $\mathbb{Z}^*$, and also that these threshold cases correspond to discontinuities of the mapping $\mu \mapsto \xi_k$ from $\Phi \to [0,1]$. Hence, this theorem provides an essentially complete solution to the size and degree distribution of the k-core, given our particular methods of asymptotic parametrization.

9.2 The k-Core and Generating Functions

In this section, we examine the solution to the k-core, as computed in Theorem 9.2.8 in more detail. Recall that, in chapter 6, we were able to express the solution to the 2-core process in terms of the fixed point of the probability generating function (p.g.f.)

$\psi_\mu(z) = \sum_{i=0}^{\infty} z^i \mu(i)$. For $k > 2$, the k-core cannot be described in quite the same way, but we may achieve a similar characterization by defining certain variants of the original p.g.f.

We begin by reviewing the expression for the p.g.f. of the 2-core residual distribution derived in chapter 6. As noted above, the solution to residual distribution for any endpoint removal process which all deterministic removals have residual degree at most d will be given by

$$\mu_\xi(i) = \nu_{(x,y)}(i) = \left(\frac{1}{1-y}\right) \sum_{j=i}^{\infty} \mu(j) \operatorname{Bin}_{(j,1-x)}(i).$$

for all $i > d$, where $x = x_\xi = e^{-w_\xi[R]}$ and $y = y_\xi = e^{-w_\xi[D]}$.

Although the $\nu_{(x,y)}(i)$ for $i \in \mathbb{Z}^*$ do not generally constitute a valid probability distribution (since they may not sum to 1), we may nevertheless define the generating function in the same way. As shown in chapter 6, this function will satisfy

$$\psi_{\nu_{(x,y)}}(z) = \sum_{i=0}^{\infty} \nu_{(x,y)}(i) z^i = \frac{\psi(x + (1-x)z)}{1-y}$$

in terms of the original p.g.f. $\psi = \psi_\mu$.

Now, in the special case where $d = 0$, we may immediately compute the p.g.f. for the actual residual distribution μ_ξ, since in this case $\nu_{(x,y)}$ and μ_ξ differ only in the coordinate $\mu_\xi(0)$. Since this coordinate corresponds to the constant term of the generating function, the p.g.f. for μ_ξ will have the form

$$\psi_{\mu_\xi} = \psi_{\nu_{(x,y)}}(z) + C$$

for a constant C (which depends on x and y). And, since the $\mu_\xi(i)$ must sum to 1, then we must have $\psi_{\mu_\xi}(1) = 1$, and therefore this constant C is given by $C = 1 - \varphi_{(x,y)}(1) = \frac{y}{y-1}$.

Hence, for the special case $d = 0$, the p.g.f. for μ_ξ is related to the original p.g.f. by the functional transformation

$$\psi_{\mu_\xi}(z) = (\alpha_{(x,y)}\psi)(z) = \frac{\psi(x + (1-x)z) - y}{1-y}.$$

We also recall that the functional transformation $\alpha_{(x,y)}$ can be understood intuitively in terms of "moving the origin" from $(0,0)$ to the point (x,y), while leaving the top right corner $(1,1)$ of the unit square fixed.

Now, for the general k-core, the deterministic removals have residual degree at most $d = k - 2$ until the k-core is found. Hence, the p.g.f. of the residual distribution μ_ξ at any time $\xi \leq \tau[k]$ will differ from $\nu_{(x,y)}$ in coordinates $0 \leq i \leq d$. It follows that the p.g.f. of the μ_ξ will be of the form form

$$\psi_{\mu_\xi}(z) = \psi_{\nu_{(x,y)}}(z) + C_0 + C_1 z + \cdots + C_d z^d.$$

And, in this case the C_i may depend not only on x and y, but also the exact solutions to the weighted measure processes $w_\xi[D_i]$ and corresponding to deterministic removals of degree i for each $i \leq d$, as well as the solution to $w_\xi[R]$ (which can be computed from the $w_\xi[D_i]$). This is because, for degrees which are subject to both random and deterministic selection, the residual distribution will generally depend on the exact times of random and deterministic selections rather than just the total weight at time ξ.

However, in cases where deterministic selections have residual degree at most d, we are generally only interested in using $\nu_{(x,y)}$ to determine $\mu_\xi(i)$ for $i > d$, since the lower degrees constitute a finite-dimensional system which can be analyzed directly using other means. Accordingly, in this section, we will introduce certain variants of the standard probability generating function which only depend on $\mu(i)$ for $i > d$, and can thus be used to algebraically ignore the lower degrees of the residual distribution.

Now, a naive approach to constructing such variants would simply be omit the first terms $\mu(i)z^i$ for $i \leq d$ from the summation $\psi_\mu(z) = \sum_i \mu(i)z^i$, but, for algebraic reasons, we adopt a slightly different approach which involves Taylor approximations. The basic definitions are given in §9.2.1, and basic properties are explored in §9.2.2.

In §9.2.3 and §9.2.4, we demonstrate how these variants of the p.g.f. can be used to compute a solution to the k-core process in an almost identical manner to the 2-core in chapter 6. Finally, in §9.2.5, we explore the connection between the k-core of a random graph $\boldsymbol{\mathcal{G}}(\mu)$, and the corresponding structure in a Galton-Watson tree, which in this case is an infinite $(k-1)$-ary subtree.

9.2.1 Taylor Approximations of Power Series

Definition 9.2.1. For any formal power series ψ, and any integer $d \geq 0$, let us denote by $(\gamma_d\psi)$ the d'th Taylor approximation of $\psi(1)$ about z, so

$$\begin{aligned}(\gamma_0\psi)(z) &= \psi(z),\\ (\gamma_1\psi)(z) &= \psi(z) + (1-z)\psi'(z),\\ (\gamma_2\psi)(z) &= \psi(z) + (1-z)\psi'(z) + \frac{(1-z)^2}{2!}\psi''(z),\end{aligned}$$

and in general

$$(\gamma_d\psi)(z) = \sum_{i=0}^{d} \frac{(1-z)^i}{i!}\psi^{(i)}(z). \tag{9.6}$$

Proposition 9.2.1. *The formal derivative of $(\gamma_d\psi)$ is given by*

$$(\gamma_d\psi)'(z) = \frac{(1-z)^d}{d!}\psi^{(d+1)}(z), \tag{9.7}$$

Proof. We compute

$$\begin{aligned}(\gamma_d\psi)'(z) &= \sum_{i=0}^{d}(1-z)^i\frac{\psi^{(i+1)}(z)}{i!} + \sum_{i=1}^{d}(-i)(1-z)^{i-1}\frac{\psi^{(i)}(z)}{i!}\\ &= (1-z)^d\frac{\psi^{(d+1)}(z)}{d!}.\end{aligned}$$

□

Corollary 9.2.2. *The power series $(\gamma_d\psi)$ and $(\gamma_{d+1}\psi)$ are related by*

$$(\gamma_{d+1}\psi)(z) = (\gamma_d\psi)(z) + \left(\frac{1-z}{d+1}\right)(\gamma_d\psi)'(z). \tag{9.8}$$

Proof. Immediate. □

In particular, note that $(\gamma_{d+1}\psi)$ is uniquely determined by $(\gamma_d\psi)$, while $(\gamma_d\psi)$ cannot be recovered from $(\gamma_{d+1}\psi)$ since the value of $(\gamma_d\psi)(0)$ is not determined by equation (9.8). Of course, the value $(\gamma_d\psi)(0)$, then we may solve this differential equation and compute $(\gamma_d\psi)(z)$; we omit this computation, but the correct formula

$$(\gamma_d\psi)(z) = (1-z)^{d+1}\left((\gamma_d\psi)(0) + \int_{y=0}^{z}\left(\frac{d+1}{(1-y)^{d+2}}\right)(\gamma_{d+1}\psi)(y)\right)$$

is not difficult to derive.

It is also not difficult to explicitly compute the coefficients of the power series expansion for $(\gamma_d\psi)$ in terms the coefficients of $\psi(z) = \sum_i a_i z^i$. For our purposes, the entire expansion is not necessary, and we will simply compute the first two terms.

Proposition 9.2.3. *For any formal power series $\psi(z) = \sum_i a_i z^i$, we have*

$$(\gamma_d\psi)(0) = \sum_{i=0}^{d} a_i,$$
$$(\gamma_d\psi)'(0) = (d+1)a_{d+1}.$$

Proof. For the constant term $(\gamma_d\psi)(0)$, note that the derivatives of $\psi(z) = \sum_i a_i z^i$ at 0 are given by then $\psi^{(i)}(0) = i!a_i$, and hence $(\gamma_d\psi)(0) = \sum_{i=0}^{d} \frac{\psi^{(i)}(0)}{i!} = \sum_{i=0}^{d} a_i$. And, by the above proposition, the first order term can be computed by $(\gamma_d\psi)'(0) = \frac{\psi^{(d+1)}(0)}{d!} = (d+1)a_{d+1}$. □

In particular, note that the term a_{d+1} from the original power series can be recovered from $(\gamma_d\psi)$. And, since $(\gamma_{d+1}\psi)$ is determined by $(\gamma_d\psi)$, then a_{d+2} can be recovered similarly from $(\gamma_d\psi)$, as can every coefficient a_i for $i > d$. In follows that, given two power series $\psi = \sum_i a_i z^i$ and $\varphi = \sum_i b_i z^i$, we have $(\gamma_d\psi) = (\gamma_d\varphi)$ if and only if

$$\sum_{i=0}^{d} a_i = \sum_{i=0}^{d} b_i$$

and $a_i = b_i$ for all $i > d$.

For a probability distribution μ on $\mathbb{Z}^*$, we must have $\sum_{i=0}^{\infty} \mu(i) = 1$, and hence $\sum_{i=0}^{d} \mu(i)$ can be determined from the values $\mu(i)$ for $i > d$. This observation yields the following corollary.

Corollary 9.2.4. *For any distributions μ_0, μ_1, we have $(\gamma_d\mu_0) = (\gamma_d\mu_1)$ if and only if $\mu_0(i) = \mu_1(i)$ for all $i > d$.*

Proof. Immediate. □

Hence, the power series $(\gamma_d\psi_\mu)(z)$ accomplishes our objective, which was to characterize the distribution μ in such a way that does not depend on the values $\mu(i)$ for $i < d$. Moreover, since

$$(\gamma_d\psi)(0) = \sum_{i=0}^{d} \mu(i),$$

then, intuitively, if μ is the distribution of residual degrees of a given endpoint partition, all endpoints of residual degrees $j \leq d$ are "lumped together" by the function $(\gamma_d\psi)$.

We can immediately appreciate the relevance of these p.g.f. variants to the k-core, since in this case any endpoint of residual degree less than $k-1$ (i.e. true degree less than k) does not belong to the k-core, regardless of its actual residual degree. In this situation, it is appropriate (and in fact preferable) to lump these endpoint together, and only consider the total number of such endpoints, which is given by $(\gamma_{k-2}\psi_\mu)(0)$ as shown above.

9.2.2 Basic Properties of $(\gamma_d\psi_\mu)$

The transformations γ_d are valid for any formal power series, or indeed for any function which is d times differentiable, but we are specifically interested in the case where $\psi = \psi_\mu$ is the generating function for a probability distribution μ on $\mathbb{Z}^*$. In section 5.4 of chapter 5, we discussed some basic properties of the p.g.f. ψ_μ of an arbitrary distribution μ; for example, $\psi_\mu(z)$ is uniformly convergent on $(-1, 1)$, has all non-negative derivatives on the interval $[0, 1)$, and satisfies $\psi_\mu(0) \geq 0$ and $\psi_\mu(1) = 1$.

The variants $(\gamma_d\psi)$ of a p.g.f. $\psi = \psi_\mu$ satisfy certain similar, though not identical properties. For example, it is evident that $(\gamma_d\psi)$ has the same radius of convergence as ψ, since each of the derivatives $\psi^{(i)}(z)$ also have this same radius of convergence. But, as noted above, while ψ_μ completely determines the entire distribution μ, $(\gamma_d\psi_\mu)$ only determines the weights $\mu(i)$ for $i > d$.

Some additional simple properties of $(\gamma_d\psi)$ are listed below.

Proposition 9.2.5. *For any probability generating function $\psi = \psi_\mu$ and any d:*

1. $(\gamma_d\psi)(0) \geq 0$ *and* $(\gamma_d\psi(1)) = 1$;

2. *if* $d > c$ *then* $(\gamma_d\psi)(z) \geq (\gamma_c\psi)(z)$ *for* $z \in [0,1]$*;*

3. $(\gamma_d\psi)'(z) \geq 0$ *for* $z \in [0,1)$*;*

4. *if* $M(\mu) = \psi'(1) < \infty$*, then* $(\gamma_d\psi)'(1) = 0$ *for* $d > 1$*.*

Proof. For the first claim, clearly $(\gamma_d\psi)(0) = \sum_{i=0}^{d} \mu(i) > 0$, Also, since ψ has non-negative derivatives of all orders then the Taylor approximation of $\psi(1)$ about $z < 1$ must underestimate $\psi(1) = 1$, and hence $(\gamma_d\psi)(z) \leq \psi(1) = 1$.

The second claim is immediate, since each term $\frac{(1-z)^i}{i!}\psi^{(i)}(z)$ is non-negative, and the third claim claim follows from proposition 9.2.1, since $(\gamma_d\psi)'(z) = \frac{(1-z)^d\psi^{(d+1)}(z)}{d!} \geq 0$.

For the fourth claim, note that if $\psi'(1) < \infty$ then $\lim_{z\to 1}(1-z)\psi''(z) = 0$, and more generally

$$\lim_{z\to 1}(1-z)^d\psi^{(d+1)}(z) = 0,$$

and hence $(\gamma_d\psi)'(1) = 0$. □

9.2.3 Random Endpoint Removal Revisited

We now show how the Taylor approximations of the p.g.f. can be used to analyze endpoint removal processes for which the maximum residual degree of deterministic removals is bounded. We begin be recalling that in the 2-core algorithm, or indeed in any removal process for which deterministic removals all have residual degree 0, the residual distribution at time ξ is generated by

$$(\alpha_{(x_\xi,y_\xi)}\psi)(z) = \frac{\psi(x_\xi + (1-x_\xi)z) - y_\xi}{1-y_\xi},$$

where $x_\xi = 1 - e^{-w_\xi[R]}$, $y_\xi = 1 - e^{-w_\xi[D]}$.

Moreover, since the exact values of $\mu_\xi(i)$ for $i \geq 1$ are given by

$$\mu_\xi(i) = \nu_{(x_\xi,y_\xi)}(i)$$

as discussed above. Now, in the case where all deterministic removals have degree at most d, then we still have $\mu_\xi(i) = \nu_{x_\xi,y_\xi}(i)$ for $i > d$. And, since the function $(\gamma_d\psi_\xi)$ only depends

the values $\mu_\xi(i)$ for $i > d$, if follows that the solution will satisfy

$$(\gamma_d \psi_\xi)(z) = (\gamma_d \alpha_{(x_\xi, y_\xi)} \psi)(z)$$

We now have the following proposition.

Proposition 9.2.6. *For any formal power series ψ, the transformations $\psi \mapsto (\alpha_{(x,y)}\psi)$ and $\psi \mapsto (\gamma_d \psi)$ commute, so we have*

$$(\gamma_d(\alpha_{(x,y)}\psi))(z) = (\alpha_{(x,y)}(\gamma_d \psi))(z).$$

Proof. Note that derivative of the function $(\alpha_{(x,y)}\psi)(z) = \frac{\psi(z+x-x\cdot z)-y}{1-y}$ is given by

$$(\alpha_{(x,y)}\psi)'(z) = \frac{(1-x)\psi'(x+(1-x)z)}{1-y} = (1-x)\left((\alpha_{(x,y)}(\psi'))(z) + \frac{y}{1-y}\right),$$

and in general

$$(\alpha_{(x,y)}\psi)^{(i)}(z) = (1-x)^i \left((\alpha_{(x,y)}(\psi^{(i)}))(z) + \frac{y}{1-y}\right).$$

Hence

$$\begin{aligned}
(\gamma_d(\alpha_{(x,y)}\psi)) &= \sum_{i=0}^{d-1} (1-z)^i \frac{(\alpha_{(x,y)}\psi)^{(i)}(z)}{i!} \\
&= (\alpha_{(x,y)}\psi)(z) + \sum_{i=1}^{d-1} \frac{(1-z)^i (1-x)^i}{i!} \left((\alpha_{(x,y)}(\psi^{(i)}))(z) + \frac{y}{1-y}\right) \\
&= \sum_{i=0}^{d-1} \frac{(1-(x+(1-x)z))^i}{i!} \left(\frac{\psi^{(i)}(x+(1-x)z)}{1-y}\right) - \frac{y}{1-y} \\
&= (\alpha_{(x,y)}(\gamma_d \psi))(z).
\end{aligned}$$

□

It follows that we may in fact express the p.g.f. variant of a removal process as

$$(\gamma_d \psi_\xi)(z) = (\gamma_d \alpha_{(x_\xi, y_\xi)} \psi)(z) = (\alpha_{(x_\xi, y_\xi)} \gamma_d \psi)(z),$$

as stated in the following corollary.

Corollary 9.2.7. *For any endpoint removal process such that all deterministic removals in an interval have residual degree at most d:*

1. *every w.e.h.p. solution to μ_ξ satisfies*

$$(\gamma_d \psi_{\mu_\xi})(z) = (\alpha_{(x_\xi, y_\xi)} \gamma_d \psi)(z),$$

   where $x_\xi = 1 - e^{-w_\xi[R]}$, $y_\xi = 1 - e^{-w_\xi[D]}$, and $\psi = \psi_\mu$ is the p.g.f. of the initial distribution;

2. *the fraction of endpoints of residual degree at most d asymptotically satisfies*

$$\sum_{i=0}^{d} \mu_\xi(i) = (\gamma_d \psi_\xi)(0) = \frac{(\gamma_d \psi)(x_\xi) - y_\xi}{1 - y_\xi}.$$

Proof. Immediate. □

9.2.4 A Visual Solution to the k-Core Process

As for the 2-core, the solution to the k-core process can be understood visually by plotting the relevant p.g.f. variants. In this case, deterministic removals have degree at most $k-2$ (i.e. residual degree at most $k-1$) until the time $\boldsymbol{\tau}[k]$ when the k-core is found. Hence, the relevant p.g.f. variant for the k-core part of the process is the $(\gamma_{k-2}\psi)$, and we may restate Theorem 9.2.8 in terms of generating functions as follows.

Theorem 9.2.8. *Let μ be a distribution with p.g.f. $\psi_\mu = \psi$, and for any $d \geq 0$, let z_d denote the smallest fixed point of the function $(\gamma_d \psi)$ in the interval $[0, 1]$. Then, for any $k \geq 2$, the following hold w.e.h.p. for a random graph $\boldsymbol{\mathcal{G}}(\mu)$:*

1. *if $(\gamma_{k-2}\psi)'(z_{k-2}) > 1$, then the k-core contains $(1 - z_{k-2})^2 m \pm o(m)$ endpoints;*

2. *if, in addition, $z_{k-2} < 1$, then the residual distribution of the k-core converges to the unique distribution with p.g.f. φ satisfying*

$$(\gamma_{k-2}\varphi)(z) = (\alpha_{(z_{k-2}, z_{k-2})} \gamma_{k-2} \psi)(z).$$

Proof. By the corollary above, we have $(\gamma_{k-2}\psi_{\mu_\xi})(z) = (\alpha_{(x_\xi,y_\xi)}\gamma_d\psi)(z)$ for any $\xi < \tau[k]$, and since the fraction of of endpoints of degree $k-1$ or less (i.e. residual degree $k-2$ or less) is

$$(\gamma_{k-2}\psi_\xi)(0) = \frac{\psi(x_\xi) - y_\xi}{1 - y_\xi}.$$

Since random and deterministic steps alternate, we have $x_\xi = y_\xi = 1-\sqrt{1-\xi}$, and therefore any solution to $\tau[k]$ must be such that $x_{\tau[k]} = y_{\tau[k]}$ is a fixed point in the function $(\gamma_{k-2}\psi)$.

The condition $(\gamma_{k-2}\psi)'(z_{k-2}) > 1$ ensures that $(\gamma_d\psi)(z) > z$ for $z \in (z_{k-2}, z_{k-2}+\epsilon)$, so in this case the only possible solution corresponds to the smallest fixed point, and therefore satisfies $x_{\tau[k]} = y_{\tau[k]} = z_{k-2}$. It follows that $\tau[k] = 1-(1-z_{k-2})^2$, and the size of the k-core is $(1-\tau[k])m \pm o(m)$.

Also, if $z_{k-2} < 1$, the the k-core contains $\Omega(m)$ endpoints, and in this case the residual distribution is uniquely determined by the function $(\gamma_{k-2}\psi_{\tau[k]})(z)$, since we have $\mu_{\tau[k]}(i) = 0$ for $i \leq k-1$. □

The solution to a core process can thus be understood by plotting all of the p.g.f. variants in at once as in figure 9.1. Since $(\gamma_{d+1}\psi)(z) \geq \gamma_d\psi(z))$, then the lowest of the curves plotted is just the original p.g.f, followed by $(\gamma_1\psi)(z)$ and so on. And, the sequence of fixed points in each of these plots corresponds the the sequence of solutions to $\tau[k]$.

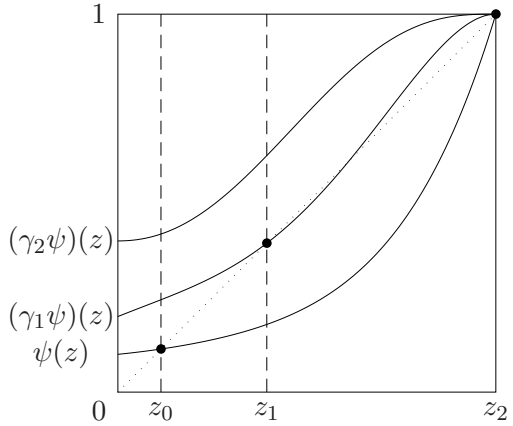


Figure 9.1: The functions $\psi(z)$, $(\gamma_1\psi)(z)$ and $(\gamma_2\psi)(z)$ for a typical distribution μ, along with the fixed points z_0, z_1, z_2. For this distribution, the graph $\boldsymbol{\mathcal{G}}(\mu)$ has large 2- and 3-cores, while the 4-core is contains at most $o(m)$ endpoints.

9.2.5 The k-Core and Galton-Watson Trees

In chapter 7, we showed that the decomposition of a random graph in terms of its 2-core and giant component is naturally analogous to the decomposition of a Galton-Watson (GW) tree based on survival and extinction. The k-core of a random graph can similarly be related to an infinite $(k-1)$-ary sub-tree of a GW tree, the existence and structure of which can be described in terms of the p.g.f. variants described above.

We first briefly recall the connection between the 2-core and survival of a GW tree. The extinction probability of a GW tree can be computed recursively in terms of the p.g.f. as follows. A GW tree is extinct (i.e. finite) if and only if every child of the root vertex produces a finite sub-tree. Since each of these sub-trees is independent, then the extinction probability must satisfy

$$p = \sum_{i=0}^{\infty} p^i = \psi_\mu(p).$$

Hence, the extinction probability must be a fixed point of the p.g.f. In order to choose the correct fixed point, we may take the limit as $t \to \infty$ of the probability of extinction after t generations, which converges from below to the smallest fixed point z_μ. And, it follows that the survival probability is $1 - z_\mu$.

Now, survival of a branching process is not directly related to the 2-core of a random graph. But, intuitively, the event that a given edge belongs to the 2-core of $\boldsymbol{\mathcal{G}}(\mu)$ corresponds to survival of the corresponding GW tree (i.e. the search tree) in both directions, and this leads to the solution of the size of the 2-core, in terms of the fraction of the original endpoint (or edge) set, as $(1 - z_\mu)^2$.

By similar analogy, the k-core corresponds to the event that a GW tree has an infinite $(k-1)$-ary sub-tree. In this this case, an edge of $\boldsymbol{\mathcal{G}}(\mu)$ belongs to the k-core if the search trees in both directions have "infinite" $(k-1)$-ary sub-trees, since each vertex in this such a tree will have degree k, including the edge connecting to its parent.

We may compute the probability of finding an infinite $(k-1)$-ary sub-tree in a GW tree similarly to the survival probability. Specifically, let us say a GW tree is *d-extinct* if

the root vertex has at most d children which are not d-extinct. Hence, 0-extinction is just ordinary extinction as defined above, and d-extinction is the complement of containing an infinite $(d+1)$-ary sub-tree (with the same root vertex). The probability of d-extinction must therefore satisfy

$$\begin{aligned} p = \sum_{j=0}^{\infty} \mu(j) \sum_{i=0}^{d} \mathrm{Bin}_{j,1-p}(i) &= \sum_{i=0}^{d} \sum_{j=i}^{\infty} \binom{j}{i} (1-p)^i p^{j-i} \mu(j) \\ &= \sum_{i=0}^{d} \frac{(1-p)^i}{i!} \sum_{j=i}^{\infty} (j)_i p^{j-i} \mu(j) \\ &= \sum_{i=0}^{d} \frac{(1-p)^i}{i!} \psi^{(i)}(p) \\ &= (\gamma_d \psi)(p). \end{aligned}$$

It follows that the probability of d-extinction is a fixed point in $(\gamma_d \psi)(p)$. And, it easy to verify that the correct fixed point in this case is also the smallest fixed point z_d in $(\gamma_d \psi)$, since once again the probability of d-extinction after t generations approaches z_d from below as $t \to \infty$.

As noted above, d-extinction is equivalent to not containing an infinite $(d+1)$-ary sub-tree, which is in turn related to the $(d+2)$-core of a random graph. That is, for an edge to belong to the k-core of a random graph, there must be "infinite" $(k-2)$-ary trees descending in both directions from v. Hence, the above computation corresponds to the fraction

$$(1 - z_{k-2})^2$$

of endpoints in a random graph which belong to the k-core.

9.3 Termination of the k-Core Process

For $k \geq 3$, it is possible to demonstrate that the k-core must be either "giant" (i.e. of size $\Omega(1)$) or empty w.p.h.p. While there are perhaps many ways to accomplish this, we shall make use of the results from section 6.3 from chapter 6 regarding sub-critical graphs.

We begin by demonstrating that the exact residual degrees of endpoints below $k-1$ do not affect the existence or size of the k-core of a random graph.

Proposition 9.3.1. *Let $H = (A, V)$ be an even endpoint partition, and let $H' = (A', V')$ be a second endpoint partition such that:*

1. $|A'| = |A|$;
2. $\mu_{H'}(i) = \mu_H(i)$ *for all* $i > k-2$;
3. $\mu_{H'}(0) = \sum_{i=0}^{k-2} \mu_H(i)$, *and therefore* $\mu_{H'}(i) = 0$ *for all* $0 < i < k-1$.

Then the number of endpoints in the k-cores of $\boldsymbol{\mathcal{G}}(A, V)$ and $\boldsymbol{\mathcal{G}}(A', V')$ are identically distributed.

Proof. It suffices to show that the stopping time $\boldsymbol{\tau}[k]$ of the core algorithm is identically distributed for wither initial endpoint partition H or H'. To prove this, note that for any $t < \boldsymbol{\tau}[k]$, and any $i > k-2$, endpoints of residual degree i are only subject to random selection, and it follows that for any $t \leq \boldsymbol{\tau}[k]$, the distribution of the random variables $\boldsymbol{\mu}_t(i)$ for $i > k-2$ depend only on the initial values of $\mu_0(i)$ for $i > k-2$. In this case, by assumption, these values are identical. And, $\boldsymbol{\tau}[k]$ occurs at the first even time step when $\sum_{i>k-2} \boldsymbol{\mu}_t(i) = 1$, hence $\boldsymbol{\tau}[k]$ is identically distributed for either starting condition. □

Using the results regarding sub-critical graphs in chapter 6, we may now establish sufficient conditions for the k-core of $\boldsymbol{\mathcal{G}}(\mu)$ to be empty.

Corollary 9.3.2. *For $k > 2$, let μ be any distribution such that*

$$\sum_{i>k-2} i\mu(i) < 1.$$

Then, for any random graph $\mathbf{G} = \boldsymbol{\mathcal{G}}(\mu)$ satisfying $M(\mu_{\mathbf{G}}) \to M(\mu)$, and for which the maximum degree is $m^{1/8-\Omega(1)}$, the k-core is empty w.p.h.p.

Proof. By the above proposition, the probability that $\mathbf{G}$ has an empty k-core is the same as for a graph with limiting residual distribution μ', where

$$\mu'(0) = \sum_{i=0}^{k-2} \mu(i)$$

and $\mu'(i) = \mu(i)$ for $i > k-2$. Note that this distribution will satisfy

$$M(\mu') = \sum_{i=0}^{\infty} i\mu'(i) = \sum_{i>k-2} i\mu(i) < 1.$$

By proposition 6.3.2 from chapter 6, every connected component of this random graph contains at most one cycle w.p.h.p. Hence, clearly, the k-core for $k > 2$ of such a graph must be empty. □

Theorem 9.3.3. *Let $k > 2$ and let μ be a distribution such that $M(\mu) < \infty$, and*

$$(\gamma_{k-2}\psi_\mu)(z) > z$$

for all $z \in [0, 1)$. Then, for any random graph $\mathbf{G} = \boldsymbol{\mathcal{G}}(\mu)$ satisfying $M(\mu_{\mathbf{G}}) \to M(\mu)$, and for which the maximum degree is $m^{1/8-\Omega(1)}$, the k-core is empty w.p.h.p.

Proof. As shown above, the unique w.e.h.p. solution to the stopping time $\boldsymbol{\tau}[k]$ for such a distribution is given by $\tau[k] = 1$. It follows that the k-core contains at most $o(m)$ endpoints w.e.h.p.

To show that the k-core is actually empty, we choose an arbitrarily small but fixed $\epsilon > 0$, and consider the state of the core process at time $\xi = (1-\epsilon)^2$. The distribution at this time will satisfy

$$(\gamma_{k-2}\psi_\xi)(z) = (\alpha_{(x_\xi, y_\xi)}\psi)(z),$$

and therefore we have

$$\mu_\xi(i) = \nu_\xi(i) = \left(\frac{1}{1-y_\xi}\right) \sum_{j=i}^{\infty} \mu(j)\, \mathrm{Bin}_{(j,1-x_\xi)}(i)$$

for all $i > k-2$, where $x_\xi = y_\xi = \sqrt{1-\xi} = \epsilon$.

It is easy to see that

$$\begin{aligned}\sum_{i=0}^{\infty} i\nu_\xi(i) = \epsilon^{-1}\sum_{i=0}^{\infty} i\sum_{j=i}^{\infty}\mu(j)\,\mathrm{Bin}_{(j,\epsilon)}(i) &= \epsilon^{-1}\sum_{j=0}^{\infty}\mu(j)\sum_{i=0}^{j} i\,\mathrm{Bin}_{(j,\epsilon)}(i)\\ &= \epsilon^{-1}\sum_{j=0}^{\infty}\mu(j)j\cdot\epsilon\\ &= M(\mu).\end{aligned}$$

Also,

$$\begin{aligned}\nu_\xi(i) = \epsilon^{-1}\sum_{j=1}^{\infty}\mu(j)\,\mathrm{Bin}_{(j,\epsilon)}(i) &= \epsilon^{-1}\sum_{j=1}^{\infty}\mu(j)j\epsilon(1-\epsilon)^{j-1}\\ &= \sum_{j=1}^{\infty}\mu(j)j(1-\epsilon)^{j-1}\\ &= \psi'_\mu(1-\epsilon).\end{aligned}$$

And since $\psi'_\mu(1) = M(\mu)$, then by choosing ϵ sufficiently small, we can ensure that $M(\mu) - \nu_\xi(i) < 1$. Since all of the $\nu_\xi(i)$ are non-negative, this implies that

$$\sum_{i>k-2} i\mu_\xi(i) = \sum_{i>k-2} i\nu_\xi(i) \le M(\mu) - \nu_\xi(i) < 1.$$

Also, since $\epsilon > 0$ is small but constant, then the number of remaining endpoints at time ξ is $\Omega(m)$, and therefore the maximum degree is still $m^{1/8-\Omega(1)}$. Hence, the random graph $\boldsymbol{\mathcal{G}}(\mathbf{A}_{\lfloor\xi m\rfloor}, V)$ with limiting residual distribution μ_ξ w.p.h.p. has an empty k-core, and it follows that $\boldsymbol{\tau}[k] = m$ w.p.h.p., which implies that the original graph has an empty k-core w.p.h.p. as well. □

Part IV

Heuristics

Chapter 10

k-Orientability

A graph is *k-orientable* if its edges can be directed such that the maximum in-degree of any vertex is at most k. In this chapter we examine the k-orientability of the random graph $\boldsymbol{\mathcal{G}}(\mu)$ by tracing the execution of a simple k-orientation heuristic on a random input graph.

Unlike the algorithms analyzed in the previous chapters, this k-orientation heuristic is not guaranteed to succeed on an arbitrary input graph. Accordingly, this analysis will only yield one-sided results about the k-orientability of the random graph $\boldsymbol{\mathcal{G}}(\mu)$. Nevertheless, for a certain "well-behaved" class of distributions, the sufficient conditions for our k-orientation heuristic to succeed coincide with previously known (and relatively straightforward) necessary conditions for $\boldsymbol{\mathcal{G}}(\mu)$ to be k-orientable. Hence, for this particular restricted class of distributions, we are able to resolve the k-orientability question in both directions.

This restricted class of distributions includes the Poisson distribution, which corresponds to the the Erdős-Rényi random graph $\mathbf{G}_{n,p}$. The k-orientability problem for $\mathbf{G}_{n,p}$ is of particular interest, due to various algorithmic applications related to balanced allocation problems. Our analysis demonstrates that the k-orientability property for $\mathbf{G}_{n,p}$ exhibits a sharp threshold of the form

$$\begin{aligned} c_k &= \sup\{c : \mathbf{G}_{n,c/n} \text{ is } k\text{-orientable w.p.h.p.}\} \\ &= \inf\{c : \mathbf{G}_{n,c/n} \text{ is not } k\text{-orientable w.e.h.p.}\}. \end{aligned} \tag{10.1}$$

These sharp thresholds in turn yield precise formulas for the performance of algorithms which are based on finding a k-orientation of $\mathbf{G}_{n,p}$.

Chapter Organization

In section 10.1, we introduce the k-orientability problem in more detail, and we discuss some of the algorithmic applications of the k-orientability of the Erdős-Rényi random graph $\mathbf{G}_{n,p}$. Then, in section 10.2, we introduce our k-orientation, which involves a procedure called *excess degree reduction*.

The analysis of the k-orientation heuristic for a random input graph $\boldsymbol{\mathcal{G}}(\mu)$ appears in the last four sections of the chapter. First, in section 10.3, we derive the basic system of differential equations which govern the execution of this heuristic on a random input graph $\boldsymbol{\mathcal{G}}(\mu)$.

However, unlike in previous chapters, we are we are not able to solve these equations in general form and derive simple and algebraically concise conditions which determine whether or not the heuristic will succeed on a general random graph $\boldsymbol{\mathcal{G}}(\mu)$. Of course, for any particular distribution μ, we may in principle attempt to solve these equations explicitly, in section 10.4 we discuss certain techniques, based on the p.g.f. variants defined in the previous section, which can be used to compute such a solution.

In section 10.5, we demonstrate that, if the residual distribution μ satisfies certain regularity conditions, then the fact that our k-orientation heuristic succeeds w.p.h.p. can be ascertained algebraically without explicitly solving differential equations. These algebraic conditions are based solely on properties of the $(k+1)$-core of $\boldsymbol{\mathcal{G}}(\mu)$, and thus can be verified directly using the results from the previous chapter. And, as noted above, for this particular class of distributions, the success of our heuristic coincides with previously known conditions which are necessary in order for $\boldsymbol{\mathcal{G}}(\mu)$ to be k-orientable.

Finally, in section 10.6, we show that the Poisson distribution satisfies our regularity conditions, and this fact allows us to determine the k-orientability thresholds for $\mathbf{G}_{n,p}$.

History and Background

Due to the algorithmic applications, the k-orientability problem for $\mathbf{G}_{n,p}$ has received a certain amount of research attention. Until recently, the best known bounds were based

directly on k-core results; improved lower bounds appeared in a 2004 Ph.D. thesis [51].

The algorithm in this chapter, and the solution for $\mathbf{G}_{n,p}$ appeared in [34]; the solution for general degree distributions using the p.g.f. appeared in [31]. The k-orientability thresholds for $\mathbf{G}_{n,p}$ were independently computed by Cain, Sanders, and Wormald [18] using a different heuristic. The k-orientability thresholds computed in this chapter (and, evidently, in [18]) confirm a conjecture by Karp and Saks [44].

As mentioned, this result has many algorithmic applications, related to the family of balanced allocation problems (see e.g. [5, 23]). Specific applications include hashing [26, 58, 62], parallel storage [61] and simulation of shared memory on DMM, and concurrent accesses to parallel disks [45].

10.1 The k-Orientability Problem

As defined above, a graph is *k-orientable* if its edges can be directed in such a way that the maximum in-degree is at most k. In certain cases, the k-orientability question can be resolved immediately based on simple combinatorial arguments. For instance, since the average in-degree of any edge-orientation of a particular graph must be precisely equal to one-half of the ordinary average degree, then any graph with average degree strictly greater than $2k$ is not k-orientable. Similarly, if G contains any induced sub-graph of average degree strictly greater than $2k$, we may once again conclude that G is not k-orientable.[1]

The k-orientability question is also related to the $(k+1)$-core of a graph. Indeed, it is not difficult to see that any graph is k-orientable if and only if its $(k+1)$-core is k-orientable. This is because the $(k+1)$-core algorithm can be trivially modified to yield a simple heuristic for the k-orientability problem which simply directs edges towards vertices of degree k or less in a greedy fashion.

Accordingly, in some situations, the results on the $(k+1)$-core of $\boldsymbol{\mathcal{G}}(\mu)$ from the

[1] In fact, although it is by no means obvious, it can be shown that this condition is both necessary and sufficient[37]; that is, a graph G is k-orientable if and only if every induced sub-graph has average degree at most $2k$.

previous chapter can also answer the question of whether $\boldsymbol{\mathcal{G}}(\mu)$ is k-orientable:

- if the $(k+1)$-core is empty then $\boldsymbol{\mathcal{G}}(\mu)$ is k-orientable;
- if the $(k+1)$-core has average degree greater than $2k$, then $\boldsymbol{\mathcal{G}}(\mu)$ is not k-orientable.

Our algorithmic analysis will improve upon the first of the two bounds by establishing conditions under which a particular k-orientation heuristic succeeds despite the fact that $\boldsymbol{\mathcal{G}}(\mu)$ contains a non-empty $(k+1)$-core. Moreover, for a certain restricted class of distributions, we will demonstrate that our heuristic succeeds whenever the $(k+1)$-core of $\boldsymbol{\mathcal{G}}(\mu)$ has average degree strictly lower than $2k$, thus matching the second of these bounds.

10.1.1 k-Orientability of $\mathbf{G}_{n,p}$

By analyzing our heuristic for a Poisson distribution, we shall demonstrate that the random graph $\mathbf{G}_{n,p}$ exhibits a sharp k-orientability threshold of the form

$$\begin{aligned} c_k &= \sup\{c : \mathbf{G}_{n,c/n} \text{ is } k\text{-orientable w.p.h.p.}\} \\ &= \inf\{c : \mathbf{G}_{n,c/n} \text{ is not } k\text{-orientable w.e.h.p.}\}. \end{aligned}$$

The k-orientability threshold c_k coincides with the threshold for which that $(k+1)$-core has average degree $2k$, and therefore the values of c_k can be computed using the techniques of the previous chapter.[2]

The k-orientability problem for $\mathbf{G}_{n,p}$ is of particular interest due to certain algorithmic applications; a more comprehensive discussion of applications can be found, e.g. in [5, 23]. Applications of k-orientability generally involve variations of the following balanced allocation problem.

[2]We note that the high probability guarantee for the lower bound is weaker than for the upper bound. This is because a small number of edges can cause a graph to not be k-orientable (for instance, any graph which contains a $(2k+1)$-clique is not k-orientable). Hence, it is not possible to achieve exponentially high probability guarantee of k-orientability. On the other hand, it is not generally possible to make a graph k-orientable by removing a small number of edges, and hence the upper bound will indeed hold with exponentially high probability.

- Given m balls and n bins, each of which has a capacity of k, assume each ball is assigned 2 bins, uniformly at random. Can we place each ball to one if its assigned bins in such a way that none of the bins exceeds its capacity of k balls?

The relationship to the k-orientability problem can be understood as follows. Let us construct a graph for which the vertex set is the set of n bins; then, for each of the m balls, we create an edge connecting the two corresponding bins. This results in a random graph with exactly m edges and n vertices, which is distributed similarly (though not exactly identically) to the Erdős-Rényi random graph $\mathbf{G}_{n,m}$.[3] And, finding a solution to the balanced allocation problem described above is equivalent to finding a k-orientation of this random graph.

Determining the k-orientability threshold for $\mathbf{G}_{n,m}$ (or $\mathbf{G}_{n,p}$) will also yield a tight bound on the maximum value of m for which this problem is solvable, given k and n. Since edge corresponds to two endpoints in a graph, so the average degree will be $2m/n$, and in this case the problem will be solvable if $m < c_k n/2$ (or, more precisely, if $m = c_k n/2 - \Omega(n)$).

While the balanced allocation problem itself is somewhat abstract, there are various more concrete computational applications. We now discuss two specific applications.

Perfect Hashing

For this application, each ball corresponds to a key which must be stored in a hash table, and each bin corresponds to a memory location which is able to store k data items. Assuming the existence of two fully random hash functions, each key can therefore be assigned to two independent, uniformly random memory locations. If the balanced allocation problem is solvable, then we may accommodate m keys in a table of size $k \cdot m$, by storing each data item in one of the two locations designated by its key.

In this case, the worst-case lookup time is constant, since in order two find a particular data item, we must search a total of at most $2k$ slots, corresponding to the two possible

[3] In fact, the random graph produced in this case would be distributed identically to the random configuration used in appendix A to show that $\mathbf{G}_{n,m}$ can be simulated by the configuration model.

bins to which the key may be hashed. And, the maximum load in this case is

$$m/kn = c_k/2k,$$

since $2m/n$ is the average degree in the corresponding graph.

Parallel Storage With Redundancy

In this application, we have n parallel disks, and we wish to store a large amount of data (i.e. $\omega(n)$ data items) in such a way that an arbitrary set of m data items can be retrieved with k parallel accesses.

The scheme is as follows: each data item is stored on two uniformly random drives, again using two fully random hash functions. Then, any set of m data items will yield a graph for which the vertex set is the set of disks, and the edge set consists of the pairs of drives on which each of the m data items are stored. By finding a k-orientation of this graph, each data item can be retrieved from one of its two locations in such a way that none of the drives need to be accessed more than k times.

10.2 The Excess Degree Reduction Heuristic

In this section we present a simple recursive heuristic which attempts to find a k-orientation in a graph G. We begin with an informal overview, and then we prove the correctness of this heuristic in §10.2.1.

Given a graph $G = (A, V, E)$, and a mapping $K : V \to \mathbb{N}$, let us define a K-orientation to be an orientation of the edges of G such that the in-degree of each $v \in V$ is at most $K(v)$. For a constant k, the k-orientation problem is thus a special case of the K-orientation problem, by letting $K(v) = k$ for all k.

For an instance of the K-orientation problem, and a particular vertex $v \in V$, we will say that:

- v is *unconstrained* if $\deg(v) \leq K(v)$;

- v is *over-constrained* if $\deg(v) > 2K(v)$;
- v is *partially constrained* if $K(v) < \deg(v) \leq 2K(v)$.

Our K-orientation heuristic is recursive, in that, given a graph G, and a mapping K, the algorithm will produce a modified graph G' and a modified in-degree constraint $K' : V' \to \mathbb{N}$ such that if G' is K'-orientable, then G is K-orientable as well. The methods used to construct the graph G' are based on two observations.

First, if v is unconstrained, then every every edge incident on v can be directed toward v. In this case, we may simply choose any edge e incident on an unconstrained vertex v, direct this edge towards v, and recursively K-orient the edges in the graph $G' = G - e$.

The situation becomes non-trivial when there are no unconstrained vertices. Now, if every vertex is over-constrained, then the graph G is evidently not K-orientable. Otherwise, we choose a partially constrained vertex v and employ a procedure called *excess degree reduction*.

Note that a partially constrained vertex has the property that at least half of its incident edges may be directed inwards. Accordingly, we shall arbitrarily choose two endpoints a_1, a_2 which belong to v, with the intention of guaranteeing that at most one of a_1, a_2 are ultimately directed toward v. We consider two cases:

1. $\{a_1, a_2\}$ is an edge in G;
2. a_1 and a_2 are connected to two other endpoints b_1, b_2 in G.

In the first case, it is trivial that exactly one of $\{a_1, a_2\}$ must be directed inward. Hence we direct this edge in either direction, and then remove this edge from G to produce the graph G'. We then set $K'(v) = K(v) - 1$, noting that G is clearly K-orientable (if and) only if G' is K'-orientable.

In the second case, note that exactly one of a_1, a_2 will be directed inward if and only if exactly one of the matched endpoints b_1, b_2 is directed inward. We can ensure that this second condition occurs by connecting b_1 and b_2 with an edge. Accordingly, we remove the

endpoints a_1 and a_2 and create a new edge $\{b_1, b_2\}$ in the graph G'. And, once again, we set $K'(v) = K(v) - 1$ and recursively K'-orient the graph G'. The diagram below illustrates these two cases on two endpoints of a vertex of degree 3.

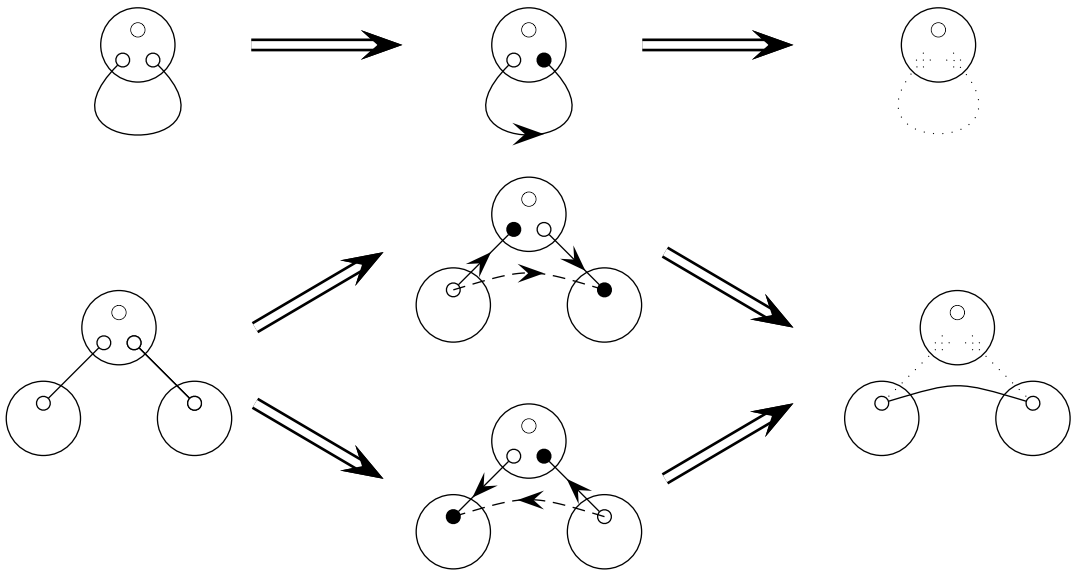


Figure 10.1: The excess degree reduction heuristic for a partially constrained vertex with $\deg(v) = 3$ and $K(v) = 2$. In the first case, we may direct the loop in either direction. In the second case, we replace two edges incident on v with a single edge as shown.

Note that, after excess degree reduction we have $\deg'(v) = \deg(v) - 2$ and $K'(v) = K(v) - 1$, and therefore after $\deg(v) - K(v)$ rounds, v will become unconstrained, in which case we may direct all incident edges towards v as discussed above. Hence, the K-orientation algorithm will proceed by repeatedly choosing a vertex v of minimum degree and applying the applicable steps as follows:

1. if v is unconstrained, direct all incident edges inwards, and remove v along with these edges;

2. if v is partially constrained, perform excess degree reduction until v is unconstrained, and then proceed as above;

3. if v is over-constrained terminate and and report failure.

10.2.1 Correctness and Conditional Uniformity

It is fairly easy to rigorously verify the correctness of the excess degree reduction heuristic described above. However, the procedure used to produce the modified graph G' which is passed to the recursive call is not quite as simple as the modifications used in our previous algorithmic analyses (e.g. the k-core or the giant component), so it is not immediately obvious that this modification preserves conditional uniformity when executed on a random input graph $\boldsymbol{\mathcal{G}}(A, V)$.

We now establish both of these facts formally; specifically, we will show that the above algorithm can be used to define an endpoint removal process which computes an upper bound on the probability that $\boldsymbol{\mathcal{G}}(A, V)$ is K-orientable for any in-degree constraint $K : V \to \mathbb{N}$. Hence, let us define, for any (even) endpoint partition (A, V), and any K, a function

$$F_K(A, V) = \mathsf{P}\left[\boldsymbol{\mathcal{G}}(A, V) \text{ is } K\text{-orientable}\right].$$

Now, if any $v \in V$ is unconstrained, meaning that $\deg(v) \leq K(v)$, then the above algorithm simply chooses an endpoint $a \in A(v)$ and removes the edge $\mathbf{E}$. This is the same procedure that we have used in all of our previous algorithms, and thus it is easy to see that conditional uniformity is maintained. Hence, we have the following proposition.

Lemma 10.2.1. *Given (A, V), choose $v \in V$ such that $\deg(v) \leq K(v)$, choose any $a_0 \in A(v)$, and choose choose $\mathbf{a}_1$ uniformly at random from $A \backslash \{a_0\}$. Then*

$$F_K(A, V) = \mathsf{E}[F_K(A \backslash \{a_0, \mathbf{a}_1\}, V)].$$

Proof. In a uniformly random matching of A, the endpoint a_0 will be matched to a uniformly random neighbor $\mathbf{a}_1$, hence we have edge $\mathbf{e}_0 = \mathbf{E}(a) \stackrel{d}{=} \{a_0, \mathbf{a}_1\}$. Let us thus define $\mathbf{A}' = A \backslash \{a_0, \mathbf{a}_1\}$, noting that the restricted matching $\mathbf{E}(\mathbf{A}') = \mathbf{E} \backslash \mathbf{e}_0$ is conditionally uniform.

Now, it is trivial that any K-orientation of $\mathbf{G} = \boldsymbol{\mathcal{G}}(A, V)$ induces a K-orientation of $\mathbf{G}' = \mathbf{G} - \mathbf{e}_0 \stackrel{d}{=} \boldsymbol{\mathcal{G}}(\mathbf{A}', V)$, and hence $F_K(A, V) \leq \mathsf{E}[F_K(\mathbf{A}', V)]$. Conversely, given a

K-orientation of $\mathbf{G}'$, we may produce an orientation of $\boldsymbol{\mathcal{G}}(A, V)$ by directing the edge $\mathbf{e}_0$ towards a_0 without changing the in-degree of any vertex other than v. Since v has degree at most $K(v)$, the in-degree of v must be at most $K(v)$ in this orientation, and therefore $F_K(A, V) \geq \mathsf{E}[F_K(\mathbf{A}', V)]$. □

In the case where v is partially constrained, the excess degree reduction heuristic uses a different method to modify the graph passed to the recursive call.

Lemma 10.2.2. *Given (A, V) and $K : V \to \mathbb{N}$, and for any $v \in V$ of degree at least 2, let K' denote the in-degree constraint with $K'(v) = K(v) - 1$, and $K'(u) = K(u)$ for $u \neq v$.*

Choose $a_0, a_1 \in A(v)$, and let $A' = A \backslash \{a_0, a_1\}$. Then

$$F_K(A, V) \geq F_{K'}(A', V).$$

Proof. Let $\mathbf{G} = \boldsymbol{\mathcal{G}}(A, V)$, and let $\mathbf{e}_0 = \mathbf{E}(a_0)$ and $\mathbf{e}_1 = \mathbf{E}(a_1)$ denote the edges containing a_0 and a_1, respectively. We now consider two cases depending on whether $\mathbf{e}_0 = \mathbf{e}_1$. First, if $\mathbf{e}_0 = \mathbf{e}_1 = \{a_0, a_1\}$, then, regardless of how this edge is directed, it will contribute 1 to the in-degree of v. In this case, $\mathbf{G} - \mathbf{e}_0$ is K'-orientable if and only if $\mathbf{G}$ is K-orientable. And, conditional on the event $\{a_0, a_1\} \in \mathbf{E}$, the restricted matching $\mathbf{E}(A')$ will be uniformly distributed and therefore

$$\mathsf{P}[\boldsymbol{\mathcal{G}}(A, V) \text{ is } K\text{-orientable} \mid \{a_0, a_1\} \in \mathbf{E}] = F_{K'}(A', V).$$

Otherwise, a_0 and a_1 match to endpoints $\mathbf{b}_0, \mathbf{b}_1 \in A'$. In this case, the edge set matching $\mathbf{E}_0 = \mathbf{E} \backslash \{\mathbf{e}_0, \mathbf{e}_1\}$ is uniformly distributed on the set of all matchings of $A' \backslash \{\mathbf{b}_0, \mathbf{b}_1\}$. Also, by symmetry, $\mathbf{b}_0, \mathbf{b}_1$ are a uniformly distributed pair of distinct endpoints in A'. Hence the edge set

$$\mathbf{E}' = \mathbf{E}_0 \cup \{\{\mathbf{b}_0, \mathbf{b}_1\}\}$$

is a uniformly random matching of A', and therefore the graph $\mathbf{G}' = (A', V, \mathbf{E}')$ is distributed identically to $\boldsymbol{\mathcal{G}}(A', V)$.

Now, assume that $\mathbf{G}'$ is K'-orientable, and w.l.o.g. assume that the edge $\{\mathbf{b}_0, \mathbf{b}_1\}$ is directed towards $\mathbf{b}_0$ in some such K'-orientation. We may produce a K-orientation of $\mathbf{G}$, as described above, by directing $\{a_0, \mathbf{b}_0\}$ towards $\mathbf{b}_0$ and $\{a_1, \mathbf{b}_1\}$ towards a_1. In doing so, we increase the in-degree of the vertex v by exactly 1, and we do not change the in-degrees of any other vertices. Hence we have a K-orientation of $\boldsymbol{\mathcal{G}}(A, V)$, and it follows that

$$\mathsf{P}[\boldsymbol{\mathcal{G}}(A, V) \textit{ is } K\textit{-orientable} \mid \{a_0, a_1\} \notin \mathbf{E}] \geq F_{K'}(A', V).$$

□

We now consider the special case where the in-degree constraint is a constant function $K(v) = k$ for all k, and we denote by $F_k(A, V)$ the probability that (A, V) is k-orientable. As noted above, for any vertex of degree $j > k$, if we perform $j - k$ rounds of excess degree reduction, both the degree of v and the in-degree constraint will drop to $j - 2(j-k) = 2k - j$, at which point v becomes unconstrained. Thus, by repeated application of lemma 10.2.2, we have the following corollary.

Corollary 10.2.3. *Let (A, V) be an endpoint partition, and let v be any vertex of degree j for $k < j \leq 2k$. Let $\{a_1, \ldots, a_{2(j-k)}\}$ be a set of $2(j-k)$ endpoints which belong to v, and let $A' = A - \{a_1, \ldots, a_{2(j-k)}\}$ Then*

$$F_k(A, V) \geq F_k(A', V).$$

Proof. Immediate. □

Based on lemma 10.2.1 and corollary 10.2.3, we have now established the correctness of the following algorithm which computes an upper bound, in expectation, on the probability $F_k(A, V)$, that $\boldsymbol{\mathcal{G}}(A, V)$ is k-orientable.

Algorithm 10.2.1 (CM k-orientability Process)**.** *Given an (even) endpoint partition (A, V), repeat the following until success or failure:*

1. *if A is empty then terminate with success; otherwise choose a vertex v of minimum positive degree;*

2. *if* $\deg(v) > 2k$*, terminate with failure;*[4]

3. *if* $\deg(v) \leq k$*, repeat the following loop until* v *is empty:*

   3a. *remove an endpoint from* v*;*

   3b. *remove an endpoint chosen uniformly at random;*

4. *if* $\deg(v) = j$ *for* $k < j \leq 2k$*, first remove* $2(j-k)$ *endpoints from* v*; then repeat the loop in step 3 until* v *is empty.*

Theorem 10.2.4. *The probability that the CM* k*-orientability process terminates with success is at most* $F_k(A, V)$*.*

Proof. Immediate. □

10.2.2 Intuitive Motivation

Before we rigorously analyze this heuristic, we give some informal justification as two why one might expect, or at least hope, that the excess degree reduction heuristic will preform reasonably well.

First, any graph (random or otherwise) is k-orientable if and only if its $(k+1)$-core is k-orientable. The behavior of the k-orientation heuristic whenever the minimum degree is k or less is therefore optimal: we simply choose an unconstrained vertex and direct all incident edges inwards. Accordingly, the quality of this, or any other k-orientation heuristic, must therefore be judged based on decisions which are made when the minimum degree is strictly greater than k.

Also, any graph with average degree greater than $2k$ is not k-orientable. Hence, even though the heuristic defined above does not fail until the *minimum* degree exceeds $2k$, we may as well fail as soon as the *average* degree exceeds $2k$.

[4]The failure condition will be discussed further in the next section; briefly, though, any graph with minimum degree $2k$ is k-orientable if and only if this graph is $2k$-regular. Moreover, our methods of asymptotic parametrization cannot distinguish a $2k$-regular graph from a graph with $o(n)$ vertices of degree greater than $2k$. Hence, from our perspective, we may as well consider $\deg(v) \geq 2k$ rather than $\deg(v) > 2k$ to be a failure condition

Now, if the average degree is at most $2k$, and the minimum degree $\deg(v) = j$ is strictly greater than k, the typical behavior of the above heuristic is as follows. First, we remove $2(j-k)$ endpoints deterministically from the vertex v, at which point the number of remaining endpoints is $j - 2(j-k) = 2k - j$. Then, we alternate deterministic and random removals until v becomes empty, and in most cases, this will result in $2k - j$ additional removals of each kind, for a total of

$$2(j-k) + 2(2k-j) = 2k$$

total endpoints removed.[5]

Since only one vertex is removed during such an iteration, then we have a ratio of $2k$ endpoint removals for each vertex removal. It is therefore evident that, provided that the average degree is at most $2k$ at the start of the iteration, the average degree will not increase at the the end of the iteration. And, if the average degree is strictly less than $2k$, then such an iteration will result in a strict decrease in the average degree.

Intuitively, then, if the average degree does ultimately increase above $2k$, this will not occur during the excess degree reduction part of the heuristic. Indeed, the only iterations which can cause an increase of the average degree are greedy iterations during which the heuristic behaves optimally.

Of course, we cannot draw any rigorous conclusions from this observation, since the fact that the average degree does not increase during excess degree reduction does not imply that this heuristic is optimal. It may well be the case that, although the average degree does not increase immediately, the modifications made during excess degree reduction indirectly cause the average to degree to increase at some later time.

Nevertheless, at least informally, the ratio of $2k$ endpoints per vertex removed suggests that this heuristic should succeed in "typical" cases. In fact, as we shall see in section 10.5, if the residual distribution μ satisfies certain regularity conditions, then a relatively simple

[5]It may be the case that the number of endpoints removed is strictly less than $2k$, but this occurs rarely, and only if one of the randomly removed endpoints belongs to v. This is discussed further in the next section.

variation of this average degree argument can be used to demonstrate that the k-orientation heuristic will succeed on the random graph $\boldsymbol{\mathcal{G}}(\mu)$, provided that the $(k+1)$-core has average degree less than $2k$.

On the other hand, the success or failure of the k-orientation heuristic cannot always be determined from the average degree of the $(k+1)$-core. Accordingly, our general analysis of this heuristic, which is described in sections 10.3 and 10.4, does not rely on this average degree argument, and instead makes use of the differential equations technique.

10.3 Differential Equations for the k-Orientability Process

In this section, we derive a system of differential equations which govern the execution of the k-orientation process defined above. We will not discuss how to actually solve these differential equations; this task is postponed to the next two sections.

In certain ways, the k-orientability process resembles the k-core process analyzed in the previous chapter. In fact, whenever the minimum positive degree is at most k, the two processes are essentially identical, since random and deterministic steps alternate, and all deterministic selections have minimum degree. And, even when the minimum degree exceeds k, the maximum degree of any deterministically chosen endpoint is at most $2k$. It follows that the values of $\mu_\xi(i)$ for values of $i \geq 2k$ (i.e. true degree $i \geq 2k+1$) can be computed using generating functions as described in the previous chapter, provided that we are able to determine the relative frequency of random and deterministic steps over time.

However, unlike for the k-core process, random and deterministic steps do not always alternate, and therefore the solutions to the processes $(\boldsymbol{\sigma}_t[R])$ and $(\boldsymbol{\sigma}_t[D])$ will in general be non-trivial. Moreover, the number of random and deterministic steps do not uniquely yield determine $\mu_\xi(i)$ for $i < 2k$, since these degrees are subject to both random and deterministic removal.

Also, at a granular level, the k-orientation algorithm is somewhat more complicated than the algorithms we have analyzed previously. This is because the behavior of duration

of a given iteration of the algorithm depends on the minimum degree at the start of the iteration. And, since the minimum degree is a discrete quantity, it cannot be scaled and represented topologically according to our standard methods.

As a result, the definition of the k-orientation process does not immediately translate to a complete system of differential equations. We shall therefore first derive a preliminary, but incomplete, set of differential constraints which can indeed be deduced relatively easily from the process definition. Then, we shall refine these preliminary constraints by ruling out certain impossible scenarios.

Overview

First, in §10.3.1, we derive some basic differential equations which describe the change in state during a single iteration of the k-orientation algorithm. Then, in §10.3.2, we discuss problems related to the minimum degree. In particular, since the true minimum degree cannot be represented topologically by scaling, we shall instead define a related quantity called the *asymptotic minimum degree*, which yields an upper bound on the true minimum degree.

In §10.3.3 we shall derive certain inequalities which restrict the possible behavior of the true minimum degree based on the asymptotic minimum degree and the residual distribution. Then, in §10.3.4, we use these inequalities to compute the ratio $\frac{d\sigma_\xi[D]}{d\sigma_\xi[R]}$ of deterministic to random steps as a function of the residual distribution μ_ξ.

At this point we will in fact have a complete system of differential equations which govern the k-orientation process. However, there are still some remaining issues which must be addressed. First, due to the discrete dependence on the asymptotic minimum degree, the formula for $\frac{\sigma_\xi[D]}{\sigma_\xi[R]}$ as a function of μ_ξ is not continuous. It is therefore not immediately clear that these differential equations are stable; that is, there may be more than one solution for a particular initial residual distribution μ. Moreover, aside from the issue of unique solvability, the k-orientation process will often undergo transitions where the asymptotic minimum degree and the ratio $\frac{d\sigma_\xi[D]}{d\sigma_\xi[R]}$ change discontinuously, and in these

cases, the behavior is not adequately described by differential equations. Hence, in §10.3.5 and §10.3.6, we investigate issues related to stability and transition points.

Finally, in §10.3.7 we discuss the termination of the k-orientation process. Since the terminal behavior depends on the last $o(m)$ steps of the algorithm, differential analysis alone cannot guarantee successful termination. However, using results about the w.p.h.p. non-existence of a k-core from the previous chapter, we are able to derive similar conditions which guarantee successful termination of k-orientation process.

We then conclude this section in §10.3.8 with a formal theorem statement which summarizes this system of differential equations and states sufficient conditions under which the k-orientation algorithm terminates successfully on a given random input graph.

10.3.1 Basic Differential Constraints

We begin by deriving some basic differential constraints which follow immediately from the definition of the k-orientation process. At a granular level, the behavior of this process is most easily described in terms of a complete executions of the main loop, rather than individual endpoint removals. We shall therefore call such a complete execution an *iteration*. An iteration consists of several consecutive *steps*, each of which corresponds to an individual endpoint removal.

As usual, we will keep track of time based on steps, rather than iterations; hence the state $\mathbf{A}_t$ denotes the endpoint set after exactly t endpoints have been removed (so $|\mathbf{A}_t| = |\mathbf{A}_0| - t = m - t$). However, we shall also keep track of the total number of iterations, using notation which will be introduced below.

The behavior of a single iteration obviously depends on the degree of the selected vertex v, which is the minimum degree at the start of the iteration. We will call this the *degree* of the iteration. Now, the algorithm terminates with failure if the maximum degree ever exceeds $2k$ at the start of an iteration. Moreover, if the minimum degree reaches $2k$, then unless the graph is precisely $2k$-regular, the heuristic will also fail. Hence, for practical purposes, we will consider the algorithm to have failed if the minimum degree either reaches

or exceeds $2k$.

It follows that the only possible iteration degrees to consider are are $0 < j < 2k$. The following proposition describes the relationship between steps and iterations based on the iteration degree.

Proposition 10.3.1. *For any $0 < j < 2k$, all but $o(m)$ iterations of degree j w.e.h.p. include:*

1. *exactly* $\min\{2j, 2k\}$ *total steps;*
2. *exactly* $\min\{j, 2k - j\}$ *random steps;*
3. *exactly one deterministic step of degree i each $1 \leq i \leq j$.*

Proof. An iteration of degree j begins with $\max\{0, 2(j - k)\}$ deterministic steps, each of which removes a single endpoint from the chosen vertex v of minimum degree $\mathbf{deg}_t(v) = j$. After that, deterministic and random steps alternate until all endpoints which belong to the vertex v have been removed.

Hence, unless an endpoint belonging to v is chosen uniformly at random during this iteration, the number of deterministic steps will be exactly equal to j. And, since the degree of v drops by 1 during each such step, then in this case we have exactly one deterministic removal of degree i for each $1 \leq i \leq j$.

Also, each of the last

$$j - \max\{0, 2(j - k)\} = \min\{j, 2k - j\}$$

deterministic steps are followed by a random step (again assuming that random removals do not affect v), so in this case we have have $\min\{j, 2k - j\}$ random steps, for a total of $\min\{2j, 2k\}$ steps of either kind.

Now, if one or more endpoints belonging to v are removed during random steps, then the total number of steps of either kind may be strictly lower than the totals computed

above. However, the probability of choosing an endpoint belonging to v during any particular random step in this iteration is at most $j/|\mathbf{A}_t| = O(1/(m-t))$, and it follows immediately that this occurs $o(m)$ times w.e.h.p. throughout the entire algorithm. □

Topological Representation by Measure Processes

We will represent the number of iterations topologically using measure processes as usual. Hence, we first define the following indicator random variables:

- $\mathbf{I}_t[L(j)] = 1$ indicates that an new iteration of degree j begins at time t;
- $\mathbf{I}_t[L] = \sum_{j=1}^{2k-1} \mathbf{I}_t[L(j)] = 1$ indicates that an new iteration begins at time t.

The corresponding measure processes

$$\boldsymbol{\sigma}_t[L(j)] = \sum_{s=1}^{t} \mathbf{I}_s[L(j)]$$

thus count the total number of iterations of degree j which have which have occurred so far (including the iteration which is in progress at time t)[6], and similarly for $\boldsymbol{\sigma}_t[L]$.

For asymptotic reasoning, these processes are canonically scaled as usual by dividing by the total duration m. Since the $o(m)$ iterations for which the above proposition fails to hold do not affect affect the scaled solutions, we may easily deduce w.e.h.p. constraints which relate the values of $\sigma_\xi[R]$, $\sigma_\xi[D(i)]$, and ξ to the $\sigma_\xi[L(j)]$ for $0 < j < 2k$. Since we are primarily interested in differential analysis, we shall state these constraints in differential form as follows.

Corollary 10.3.2. *The differential constraints*

$$\begin{aligned} d\sigma_\xi[R] &= \sum_{0<j<2k} (\min\{j, 2k-j\}) d\sigma_\xi[L(j)], \\ d\sigma_\xi[D(i)] &= \sum_{i\leq j<2k} d\sigma_\xi[L(j)], \end{aligned} \tag{10.2}$$

[6]At a technical level, the beginning of a new iteration is the first time that an endpoint is removed deterministically from a new vertex v, so the first iteration begins at time $t = 1$; and, since every iteration lasts an even number of steps, then new iterations can only begin at odd times.

and

$$d\xi = d\sigma_\xi[R] + d\sigma_\xi[D] = \sum_{0<j<2k} (\min\{2k, 2j\}) d\sigma_\xi[L(j)]$$

hold w.e.h.p. for the k-orientation process.

Proof. Immediate. □

The Unscaled Degree Distribution

We now discuss the effect of a single iteration on the state $\mathbf{A}_t$. Now, while our ultimate objective is to find a solution to the residual distribution process $(\boldsymbol{\mu}_t)$, for algebraic purposes, we will first describe the change of state in terms of the also work with the *unscaled degree distribution*, which is defined by

$$\boldsymbol{\Lambda}_t(i) = |\{v \in V : \mathbf{deg}_t(v) = i\}|.$$

Since $\boldsymbol{\Lambda}_t(i)$ is integer-valued, we will implicitly scale $\boldsymbol{\Lambda}_t(i)$ by dividing by the duration of the process, which is the initial number of endpoints $|A| = m$. Hence, a solution will satisfy

$$\boldsymbol{\Lambda}_{\lfloor \xi m \rfloor}(i)/m \to \Lambda_\xi(i)$$

The reason for dividing by the initial number of endpoints is that the measure processes $\boldsymbol{\sigma}_t[i]$ are also implicitly scaled by dividing by m. And, the increments of $\boldsymbol{\Lambda}_t(i)$ are quite simple to compute from step degrees, since $\Delta\boldsymbol{\Lambda}_t(i) = \Delta\boldsymbol{\sigma}_t[i+1] - \Delta\boldsymbol{\sigma}_t[i]$, so scaling $\boldsymbol{\Lambda}_t(i)$ in the same way yields the simple differential constraint

$$d\Lambda_\xi(i) = d\sigma_\xi[i+1] - d\sigma_\xi[i].$$

It is of course trivial to recover the residual distribution μ_ξ from Λ_ξ, since the total number of endpoints of residual degree 1 is $(i+1)\boldsymbol{\Lambda}_t(i+1)$, and therefore

$$\mu_\xi(i) = \frac{(i+1)\Lambda_\xi(i+1)}{1-\xi}.$$

Moreover, we will not abandon the residual distribution entirely, since residual distribution most concisely describes the degrees of random selections via the w.e.h.p. differential constraint

$$d\sigma_\xi[R(i)] = \mu_\xi(i-1)d\sigma_\xi[R].$$

By corollary 10.3.2, we can now express $d\Lambda_\xi(j)$ in terms of iteration degrees as follows.

Proposition 10.3.3. *Every w.e.h.p. solution to the k-orientation process satisfies*

$$d\Lambda_\xi(j) = (\mu_\xi(j) - \mu_\xi(j-1))d\sigma_\xi[R] - d\sigma_\xi[L(j)]. \tag{10.3}$$

for $j > 0$, and

$$d\Lambda_\xi(0) = \mu_\xi(0)d\sigma_\xi[R] - d\sigma_\xi[L]. \tag{10.4}$$

Proof. As noted above, the degrees of random steps satisfy $d\sigma_\xi[R(j)] = \mu_\xi(j-1)d\sigma_\xi[R]$ w.e.h.p., which accounts for the first term in the (10.3). For the second term, every iteration of degree less than j includes no deterministic steps of degree either j or $j+1$, and all but $o(m)$ iterations of degree greater than j include one deterministic step of each degree j and $j+1$, and therefore

$$d\sigma_\xi[D(j+1)] - d\sigma_\xi[D(j)] = -d\sigma_\xi[L(j)].$$

For the second equation, we note that $d\sigma_\xi[0] = 0$, and therefore $d\Lambda_\xi[0] = d\sigma_\xi[1]$. And, all but $o(m)$ iterations include one deterministic step of degree 1, while random steps of degree 1 satisfy $d\sigma_\xi[R(1)] = \mu_\xi(0)d\sigma_\xi[R]$. □

10.3.2 The Asymptotic Minimum Degree

By definition, the degree of an iteration beginning at time $t+1$ is equal to the minimum (positive) degree at time t, which we shall denote by

$$\mathbf{dmin}_t = \min\{i > 0 : \mathbf{\Lambda}_t(i) > 0\}.$$

However, the minimum degree is not asymptotically "visible," since even if $\Lambda_\xi(i) = 0$ for $0 < i < j$, we may have $o(m)$ vertices of degrees lower than j at time $t = \lfloor \xi m \rfloor$. Accordingly,

we will instead work with a related quantity called the *asymptotic minimum degree*, which is similarly defined by

$$\text{dmin}_\xi = \min\{i > 0 : \Lambda_\xi(i) > 0\}.$$

Note that the asymptotic minimum degree is a function of the *solution* to the process, and not the process itself. In other words, the asymptotic minimum degree cannot be determined from the sample path, and is only meaningful in the asymptotic limit. In the terminology of chapter 4, the condition that $\text{dmin}_\xi = j$ is a meta-event, which can be expressed as the conjunction of the convergence $\sum_{i=1}^{j-1} \frac{\mathbf{\Lambda}_{\lfloor \xi m \rfloor}(j)}{m} \to 0$ and the separation $\frac{\mathbf{\Lambda}_{\lfloor \xi m \rfloor}(j)}{m} \nrightarrow 0$.

Degrees of Iterations

As discussed above, the condition $\text{dmin}_\xi = j$ does not necessarily imply that the true minimum degree satisfies $\mathbf{dmin}_{\lfloor \xi m \rfloor} = j$, and thus the relationship between dmin_ξ and iteration degrees is non-trivial. On the other hand, we may observe that if $\text{dmin}_\xi = j$, then we must have $\frac{\mathbf{\Lambda}_{\lfloor \xi m \rfloor}(j)}{m} \to \Lambda_\xi(j) > 0$, which in turn implies that $\mathbf{\Lambda}_{\lfloor \xi m \rfloor}(j) = \Omega(m) > 0$.

Therefore, the asymptotic minimum degree is an upper bound on the true minimum degree. We may derive the following bound on iteration degrees.

Proposition 10.3.4. *The constraint $d\sigma_\xi[L(i)] = 0$ for $i > \text{dmin}_\xi$ holds u.a.*

Proof. As noted above if $\text{dmin}_\xi = j$ then $\mathbf{\Lambda}_{\lfloor \xi m \rfloor}(j) = \Omega(m)$; and, since the increments of the unscaled distribution are bounded by $|\Delta \mathbf{\Lambda}_t(j)| \leq 1$ (always), then in this case we must also have $\mathbf{\Lambda}_t(j) = \Omega(m)$ for an entire neighborhood $t/m \in (\xi - \epsilon, \xi + \epsilon)$.[7] Hence, no iterations of degree strictly greater than $\text{dmin}_\xi = j$ can take place in this neighborhood, and we have $d\sigma_\xi[L(i)] = 0$ for $i > j$. □

[7] Indeed, perhaps the closest analogy to the asymptotic minimum degree which can actually be observed from the sample path is the maximum value of $\mathbf{dmin}_t$ for t in a small neighborhood of $\lfloor \xi m \rfloor$. However, this is not quite accurate, since it may be the case that $\text{dmin}_\xi = j$ and yet $\mathbf{dmin}_t < j$ for all $t \in (\xi - \epsilon, \xi + \epsilon)$Accordingly, it is best to think of dmin_ξ as simply a property of the solution which is only meaningful asymptotically.

Of course, this equation does not specify the exact values of $d\sigma_\xi[L(j)]/d\xi$ for all j. However, as we shall see, these derivatives are in fact uniquely determined by this upper bound, along with the constraints derived above in corollary 10.3.2 and proposition 10.3.3.

The Continuation of (dmin_ξ)

Since (dmin_ξ) is a discrete function, the derivative is not a particularly useful tool reasoning about its behavior in a small neighborhood of a particular point ξ. Hence, we shall instead work with the *continuation* for this purpose; as defined in chapter 4, the continuation of the function (dmin_ξ) at a point ξ, which is denoted by

$$\mathrm{dmin}_{\xi^+}$$

is the equivalence class of functions containing (dmin_ξ), based on the relation $(f_\xi) \equiv (g_\xi)$ if and only if there exists $\epsilon > 0$ such that $f_\zeta = g_\zeta$ for all $\zeta \in (\xi, \xi + \epsilon)$.

Accordingly, if $\mathrm{dmin}_{\xi^+} = j$ then $\mathrm{dmin}_\zeta = j$ for $j \in (\xi, \xi + \epsilon)$. Of course, it is not necessarily the case that for each ξ, there exists a constant j such that $\mathrm{dmin}_{\xi^+} = j$. Nevertheless, we may still reason about the continuation using inequalities. In particular we have the following basic proposition.

Proposition 10.3.5. *The constraint* $\mathrm{dmin}_{\xi^+} \leq \mathrm{dmin}_\xi$ *holds u.a.*

Proof. If $\mathrm{dmin}_\xi = j$ then $\Lambda_\xi(j) > 0$, and hence by continuity, $\Lambda_\zeta(j) > 0$ for $j \in (\xi, \xi+\epsilon)$. □

We may also relate continuation dmin_{ξ^+} to the differentials $d\Lambda_\xi(i)$ as follows. First, recall that, as discussed in chapter 3, we are working with *right-derivatives*, meaning that, for a function (f_ξ), the expression $\frac{df_\xi}{d\xi}$ is defined as

$$\frac{df_\xi}{d\xi} = \lim_{\epsilon \to 0^+} \frac{f_{\xi+\epsilon} - f_\xi}{\epsilon}$$

if this limit exists.

Similarly to the continuation, even if the right-derivative of a particular function does not exist, we may still reason about its differential behavior using inequalities. For example,

if $df_\xi > 0$, then the function (f_ξ) must be strictly increasing at ξ, meaning that there exists $\epsilon > 0$ such that $f_\zeta > f_\xi$ for $\zeta \in (\xi, \xi + \epsilon)$. Another important fact about the right-derivative is that it respects right-hand limits, and in particular,

$$\frac{df_\xi}{d\xi} \leq \limsup_{\zeta \to \xi^+} \frac{df_\zeta}{d\zeta}.$$

These observations allow us to derive the following basic inequalities.

Proposition 10.3.6. *The following constraints hold u.a.:*

1. *if* $d\Lambda_\xi(i) > 0$ *then* $\mathrm{dmin}_{\xi^+} \leq i$*;*
2. *if* $\mathrm{dmin}_{\xi^+} < i$ *then* $d\sigma_\xi[L(i)] = 0$*.*

Proof. For the first claim, recall that $(\Lambda_\xi(i))$ is a non-negative function, and hence if $d\Lambda_\xi(i) > 0$ then $\Lambda_\zeta(0) > \Lambda_\xi(j) \geq 0$ for $\zeta \in (\xi, \xi + \epsilon)$, and therefore by definition $\mathrm{dmin}_\zeta \leq j$ in this interval.

For the second claim, if $\mathrm{dmin}_\zeta \leq j$ holds for $\zeta \in (\xi, \xi + \epsilon)$, then for any $i > j$, we have $d\sigma_\zeta[L(i)] = 0$ in this interval. It thus follows by right-continuity that $d\sigma_\xi[L(i)] = 0$ as well. □

10.3.3 Derived Constraints and Inequalities

The differential constraints we have derived so far are relatively elementary, in that they all follow immediately either from the definition of the k-orientation process, or from basic facts about continuous functions. However, by combining these constraints, we can now derive two more powerful inequalities on the differentials $d\sigma_\xi[L(i)]$, which, as we shall see, will ultimately yield exact expressions for $\frac{d\sigma_\xi[L(i)]}{d\xi}$ as function of the residual distribution μ_ξ and the asymptotic minimum degree dmin_ξ.

We first consider a situation where $\Lambda_\xi(j) = 0$ for some given degree j. In this case, the number of vertices of degree j is $o(m)$ (but, as noted above, not necessarily 0). And, since every iteration of degree j removes a vertex of degree j, then the number of such iterations

in an interval $(\xi, \xi+\epsilon)$ is bounded above by the number of newly produced vertices of degree j.

Since the only way to produce a new vertex of degree j is via a random removal of degree $j+1$ (i.e. residual degree j), then if $\Lambda_\xi(j) = 0$, we may bound $d\sigma_\xi[L(j)]$ based on $\mu_\xi(j)$ as follows.

Proposition 10.3.7. *The constraint*

$$d\sigma_\xi[L(j)] \leq \mu_\xi(j) d\sigma_\xi[R] \quad \textit{whenever} \quad \Lambda_\xi(j) = 0, \tag{10.5}$$

holds w.e.h.p. for all $\xi \in [0,1)$.

Proof. Note that if $\Lambda_\xi(j) = 0$ then $\mu_\xi(j-1) = 0$, and therefore equation (10.3) from proposition 10.3.3 yields

$$d\Lambda_\xi(j) = \mu_\xi(j) d\sigma_\xi[R] - d\sigma_\xi[L(j)]. \tag{10.6}$$

Since $(\Lambda_\xi(j))$ is non-negative, then we must have $d\Lambda_\xi(j) \geq 0$, and therefore $d\sigma_\xi[L(j)] \leq \mu_\xi(j) d\sigma_\xi[R]$. □

While the absence of vertices of degree j limits the number of possible iterations of degree j, the presence of vertices of degree j limits the number of possible iterations of degrees greater than j. Specifically, every random endpoint removal produces a new vertex of degree j with probability $\mu_\xi(j) > 0$. And, since the degree of any iteration is equal to the (true) minimum degree $\mathbf{dmin}_t$, an iteration of degree $j+1$ cannot occur until all lower degree vertices have been removed. This observation allows us to derive an upper bound on $d\sigma_\xi[L(j+1)]$ based on $\mu_\xi(j)$.

Proposition 10.3.8. *The constraint*

$$d\sigma_\xi[L(j+1)] \leq \max\{0, d\sigma_\xi[L] - d\sigma_\xi[R]\mu_\xi(j)\} \tag{10.7}$$

holds w.e.h.p. for all $\xi \in [0,1)$.

Proof. First, note that it suffices to prove that (10.7) holds outside of a set of measure zero, since in this case, due to the continuity of the function $(\mu_\xi(j))$, this inequality will extend to all $\xi \in [0, 1)$.

Moreover, if $\Lambda_\xi(j) > 0$, then $d\sigma_\xi[L(j+1)] = 0$ by proposition 10.3.4, so (10.7) holds trivially, so it suffices to consider the case where $\Lambda_\xi(j) = 0$.

If $\Lambda_\xi(j) = 0$, then the differential $d\Lambda_\xi(j)$ will once again satisfy equation (10.6). Also, since $(\Lambda_\xi(j))$ is non-negative, then we must have $d\Lambda_\xi(j) \geq 0$. Now, the function $(\Lambda_\xi(j))$ is Lipschitz continuous and therefore almost everywhere (right-) differentiable. Hence, we may restrict our attention to values of ξ for which the right-derivative $\frac{d\Lambda_\xi(j)}{d\xi}$ exists. In this case, we may assume that either $d\Lambda_\xi(j) = 0$ or $d\Lambda_\xi(j) > 0$.

If $d\Lambda_\xi(j) > 0$ then we must have $\mathrm{dmin}_{\xi^+} \leq j$ and therefore $d\sigma_\xi[L(j+1)] = 0$ by the proposition 10.3.6. Therefore, the only remaining case to consider is if both $\Lambda_\xi(j) = 0$ and $d\Lambda_\xi(j) = 0$. In this situation, (10.6) yields the strict equality

$$d\sigma_\xi[L(j)] = d\sigma_\xi[R]\mu_\xi(j),$$

and the inequality (10.7) follows from the fact that

$$d\sigma_\xi[L] = \sum_{i=1}^{2k-1} d\sigma_\xi[L(i)] \geq d\sigma_\xi[L(j+1)] + d\sigma_\xi[L(j)].$$

□

10.3.4 Deterministic and Random Selections

Using the differential inequalities from propositions 10.3.7 and 10.3.8, we are now able to compute the frequency of deterministic and random steps as a function of the residual distribution μ_ξ. Since $d\sigma_\xi[D] + d\sigma_\xi[R] = d\xi$, then both $d\sigma_\xi[D]$ and $d\sigma_\xi[R]$ are uniquely determined by the deterministic to random step ratio, which we shall denote by

$$\theta_\xi = \frac{d\sigma_\xi[D]}{d\sigma_\xi[R]}.$$

As usual, this expression refers to the right-derivative $\theta_\xi = \lim_{\epsilon \to 0^+} \frac{\sigma_{\xi+\epsilon}[D] - \sigma_\xi[D]}{\sigma_{\xi+\epsilon}[R] - \sigma_\xi[R]}$. As we shall see, for any w.e.h.p. solution to the k-orientation process, both functions $(\sigma_\xi[D])$

and $(\sigma_\xi[R])$ will be right-differentiable for all ξ, and the ratio θ_ξ will be strictly positive and finite.

Now, by proposition 10.3.1, both $d\sigma_\xi[D]$ and $d\sigma_\xi[R]$ can be computed directly from the $d\sigma_\xi[L(i)]$, and therefore it suffices to determine each $d\sigma_\xi[L(i)]$ based on μ_ξ. As a first step, we combine the two inequalities derived above as follows.

Proposition 10.3.9. *The following hold w.e.h.p. whenever* $\mathrm{dmin}_\xi = j+1$*:*

1. $d\sigma_\xi[L(i)] = 0$ *for* $i \notin \{j, j+1\}$*;*
2. *if* $\mu_\xi(j) \geq \frac{1}{\min\{j,2k-j\}}$ *then* $d\sigma_\xi[L(j+1)] = 0$*;*
3. *if* $\mu_\xi(j) < \frac{1}{\min\{j,2k-j\}}$ *then* $d\sigma_\xi[L(j+1)] > 0$ *and*

$$d\sigma_\xi[L(j)] = \mu_\xi(j) d\sigma_\xi[R]. \tag{10.8}$$

Proof. For the first claim, note that if $\mathrm{dmin}_\xi > i+2$, then $\Lambda_\xi(i+1) = 0$, and since the residual degree is one greater than the true degree, this means that $\mu_\xi(i) = 0$ as well. In this case, the inequality (10.5) from proposition 10.3.7 immediately yields $d\sigma_\xi[L(i)] \leq \mu_\xi(i) d\sigma_\xi[R] = 0$.

Now, the fact that $d\sigma_\xi[L(i)] = 0$ implies that $d\sigma_\xi[L] = d\sigma_\xi[L(j)] + d\sigma_\xi[L(j+1)]$, and we may therefore express (10.7) from proposition 10.3.8 in terms of $d\sigma_\xi[L(j)]$ by

$$\begin{aligned} d\sigma_\xi[L(j)] &= d\sigma_\xi[L] - d\sigma_\xi[L(j+1)] \\ &\geq d\sigma_\xi[L] - \max\{0, d\sigma_\xi[L] - \mu_\xi(j) d\sigma_\xi[R]\} \\ &\geq \min\{d\sigma_\xi[L], \mu_\xi(j) d\sigma_\xi[R]\}. \end{aligned}$$

Also, since $\mathrm{dmin}_\xi = j+1$, then $\Lambda_\xi(j) = 0$, and hence proposition 10.3.7 is applicable and yields the opposite inequality $d\sigma_\xi[L(i)] \leq \mu_\xi(j) d\sigma_\xi[R]$. Hence, at this point there are only two possibilities:

1. $d\sigma_\xi[L(j)] < \mu_\xi(j) d\sigma_\xi[R]$ and $d\sigma_\xi[L(j)] = d\sigma_\xi[L]$ (in which case $d\sigma_\xi[L(j+1)] = 0$);
2. $d\sigma_\xi[L(j)] = \mu_\xi(j) d\sigma_\xi[R]$.

In order to complete the proof, we now show that only one of these two cases is possible for any given $\mu_\xi(j)$. First, as computed in proposition 10.3.1, all but $o(m)$ iterations of degree j include exactly $\min\{j, 2k-j\}$ random steps. It follows that, if $d\sigma_\xi[L(j)] = d\sigma_\xi[L]$, then

$$d\sigma_\xi[R] = (\min\{j, 2k-j\})d\sigma_\xi[L(j)].$$

In particular, if $\mu_\xi(j) \leq \frac{1}{\min\{j,2k-j\}}$, then we would have

$$\mu_\xi(j)d\sigma_\xi[R] \leq d\sigma_\xi[L(j)].$$

Hence, if $\mu_\xi(j) \leq \frac{1}{\min\{j,2k-j\}}$, the first of these two possibilities cannot occur, and thus we must have $d\sigma_\xi[L(j)] = \mu_\xi(j)d\sigma_\xi[R]$.

Moreover, even if $d\sigma_\xi[L(j)] < d\sigma_\xi[L]$, the fact that iterations of degree j include $\min\{j, 2k-j\}$ random steps still yields the inequality

$$d\sigma_\xi[R] \geq (\min\{j, 2k-j\})d\sigma_\xi[L(j)].$$

Therefore, if $\mu_\xi(j) > \frac{1}{\min\{j,2k-j\}}$, then we have

$$\mu_\xi(j)d\sigma_\xi[R] > d\sigma_\xi[L(j)],$$

and we can rule out the second of the two possibilities listed above. □

It is is now relatively straightforward to compute the ratio $\theta_\xi = \frac{d\sigma_\xi[D]}{d\sigma_\xi[R]}$ from equation (10.8).

Proposition 10.3.10. *For every w.e.h.p. solution to the k-orientation process, ratio* $\theta_\xi = \frac{d\sigma_\xi[R]}{d\sigma_\xi[D]}$ *is determined as follows:*

1. *if* $\mathrm{dmin}_\xi \leq k$ *then* $\theta_\xi = 1$*;*
2. *if* $\mathrm{dmin}_\xi = j+1 > k$ *and* $\mu_\xi(j) \geq \frac{1}{2k-j}$ *then* $\theta_\xi = \frac{j}{2k-j}$*;*
3. *if* $\mathrm{dmin}_\xi = j+1 > k$ *and* $\mu_\xi(j) \leq \frac{1}{2k-j}$*, then*

$$\theta_\xi = \frac{j+1-2k\mu_\xi(j)}{2k-j-1}. \tag{10.9}$$

Proof. First, if $\text{dmin}_\xi \leq k$, then all iterations have degree at most k, and since each such iteration includes the same number of random and deterministic steps (as in the k-core algorithm), then we have $d\sigma_\xi[D] = d\sigma_\xi[R]$, and hence $\theta_\xi = 1$.

If $\text{dmin}_\xi = j + 1 > k$ and $\mu_\xi(j) \geq \frac{1}{2k-j}$, by proposition 10.3.9, we have $d\sigma_\xi[L(j)] = d\sigma_\xi[L]$ and $d\sigma_\xi[L(j+1)] = 0$. Hence, by proposition 10.3.1, each iteration will include $2k-j$ random steps and j deterministic steps, for a ratio of $\theta_\xi = \frac{j}{2k-j}$.

Finally, in the non-trivial case where $\mu_\xi(j) \leq \frac{1}{2k-j}$ and $\text{dmin}_\xi = j + 1 > k$, first note that iterations of degree either $j + 1, j \geq k$ include $2k$ total steps, and thus we have $d\sigma_\xi[L] = d\xi/2k$. The frequency of random steps can therefore be expressed as

$$\begin{aligned} d\sigma_\xi[R] &= (2k - j - 1)d\sigma_\xi[L(j + 1)] + (2k - j)d\sigma_\xi[L(j)] \\ &= \left(\frac{2k - j - 1}{2k}\right) d\xi + d\sigma_\xi[L(j)]. \end{aligned} \tag{10.10}$$

Moreover, in this case, proposition 10.3.9 yields the equation

$$d\sigma_\xi[L(j)] = \mu_\xi(j)d\sigma_\xi[R].$$

Combined with (10.10), we therefore have

$$\frac{d\sigma_\xi[R]}{d\xi} = \frac{2k - j - 1}{(1 - \mu_\xi(j))2k}, \tag{10.11}$$

and we may easily compute

$$\begin{aligned} \theta_\xi = \frac{d\sigma_\xi[D]}{d\sigma_\xi[R]} = \frac{1 - \frac{d\sigma_\xi[R]}{d\xi}}{\frac{d\sigma_\xi[R]}{d\xi}} &= \frac{(1 - \mu_\xi(j))2k - (2k - j - 1)}{2k - j - 1} \\ &= \frac{j + 1 - 2k\mu_\xi(j)}{2k - j - 1}. \end{aligned}$$

□

10.3.5 Stability Conditions

At this point, the task of deriving a complete set of differential constraints is complete, since proposition 10.3.10 uniquely determines the derivatives $\frac{d\sigma_\xi[D]}{d\xi}$ and $\frac{d\sigma_\xi[R]}{d\xi}$ as a function

of the residual distribution μ_ξ. However, the expression for θ_ξ is a continuous function of the residual function μ_ξ, since θ_ξ depends on the asymptotic minimum degree dmin_ξ. Of course, if $\mathrm{dmin}_\xi \leq k$ then we have $\theta_\xi = 1$ regardless of the actual asymptotic minimum degree. But if $\mathrm{dmin}_\xi > k$, then a small change in μ_ξ can, in principle, result in a change in dmin_ξ which in turn can discontinuously change the value of θ_ξ.

Due to this possibility, it is not immediately evident that the system of differential equations determined above will have a unique solution. Nevertheless, if we are able to prove that $\mathrm{dmin}_\xi = j+1$ holds for some extended interval $\xi \in (\zeta_1, \zeta_2)$, then (10.9) expresses θ_ξ as a Lipschitz continuous function of $\mu_\xi(j)$. It follows that, under the assumption that dmin_ξ remains constant, these differential equations are indeed uniquely solvable in such an interval.[8]

Accordingly, if we are able to compute a solution for some sub-interval $\xi \in [0, \zeta]$, and we can demonstrate that the continuation $\mathrm{dmin}_{\zeta^+} = j$ for some fixed j, then this solution can be extended uniquely to some larger interval $\xi \in [0, \zeta + \epsilon]$. We now derive conditions which imply that dmin_{ξ^+} remains constant. First, recall that the continuation satisfies the upper bound

$$\mathrm{dmin}_{\xi^+} \leq \mathrm{dmin}_\xi \,.$$

Intuitively, this means that the asymptotic minimum degree cannot increase "unexpectedly," and we only need to guard against sudden drops in dmin_ξ. Using the bounds on $d\sigma_\xi[L(i)]$, we may deduce the following lower bound on dmin_{ξ^+}.

Proposition 10.3.11. *If* $\mathrm{dmin}_\xi > j$ *and* $\mu_\xi(j) < \frac{1}{\min\{j, 2k-j\}}$ *then* $\mathrm{dmin}_{\xi^+} > j$.

Proof. First, since the function $\xi \mapsto \mu_\xi(j)$ is continuous, there exists $\epsilon > 0$ such that $\mu_\zeta(j) < \frac{1}{\min\{j, 2k-j\}}$ for $\zeta \in [\xi, \xi + \epsilon)$. We claim that $\mathrm{dmin}_\zeta \geq j+1$ in this interval.

Note that the set

$$\mathcal{X} = \{\zeta \in [\xi, \xi + \epsilon) : \mathrm{dmin}_\zeta \leq j\}$$

[8] Since the system is infinite dimensional, the unique solvability is not quite immediate from the Lipschitz condition; however, the results from previous chapters imply that, in this particular case, the infinite-dimensionality is not an issue.

is the inverse image of the open set $(0, \infty)$ under the continuous mapping $\xi \mapsto \sum_{i=1}^{j} \Lambda_\xi(i)$, and hence $\mathcal{X}$ is an open subset of $(\xi, \xi + \epsilon)$. Moreover, since $\mathrm{dmin}_\xi > j$ by assumption, then $\xi \notin \mathcal{X}$, and therefore $\mathcal{X} \subseteq (\xi, \xi + \epsilon)$.

It follows that $\mathcal{X}$ can be expressed as the disjoint union of open intervals, and therefore, unless $\mathcal{X}$ is empty, there must exist some $(\xi_0, \xi_1) \subseteq \mathcal{X}$ for which $\xi_0 \notin \mathcal{X}$ and therefore $\mathrm{dmin}_{\xi_0} > j$.

In this case, we have $\mathrm{dmin}_{\xi_0^+} \leq j$, and hence by right-continuity $d\sigma_{\xi_0}[L(i)] = 0$ for $i > j$. Moreover, since $\mathrm{dmin}_{\xi_0} > j$, then $d\sigma_{\xi_0}[L(i)]$ for $i < j \leq \mathrm{dmin}_{\xi_0} - 1$. Hence, the only possible value of i such that $d\sigma_{\xi_0}[L(i)] > 0$ is $i = j$, and thus we have $d\sigma_{\xi_0}[L] = d\sigma_{\xi_0}[L(j)]$.

This immediately implies that dmin_{ξ_0} must be exactly equal to $j+1$ since otherwise we would have $d\sigma_{\xi_0}[L(j)] = 0$ as well. However, by assumption, we have $\mu_{\xi_0}(j) < \frac{1}{\min\{j, 2k-j\}}$, and proposition 10.3.9 thus implies that $d\sigma_{\xi_0}[L(j+1)] > 0$. This is a contradiction, and therefore such a ξ_0 cannot exist in $[\xi, \xi + \epsilon)$, and it follows that $\mathcal{X}$ must be empty. □

This lower bound, combined with the fact that $\mathrm{dmin}_{\xi^+} \leq \mathrm{dmin}_\xi$ for all ξ, yields the following stability condition for the asymptotic minimum degree.

Corollary 10.3.12. *If* $\mathrm{dmin}_\xi = j + 1$ *and* $\mu_\xi(j) < \frac{1}{\min\{j, 2k-j\}}$ *then* $\mathrm{dmin}_{\xi^+} = j + 1$.

Proof. Immediate. □

10.3.6 Transitions

Based on the stability conditions derived above, a qualitative picture of the high-level behavior of the k-orientation process begins to emerge. We shall characterize this behavior as follows:

- a *stable interval* is an open interval $\xi \in (\zeta_1, \zeta_2)$ during which dmin_ξ remains constant;
- a *transition point* is any ξ which does not belong to a stable interval (i.e. a discontinuity in the function (dmin_ξ).

As discussed above, the ratio the ratio $\theta_\xi = \frac{d\sigma_\xi[R]}{d\sigma_\xi[D]}$ depends continuously on μ_ξ in any stable interval, and thus these differential equations are uniquely solvable in such in interval. Moreover, if $\mathrm{dmin}_\xi \leq k$ holds for $\xi \in (\zeta_1, \zeta_2)$, then $\theta_\xi = 1$ will hold for this entire interval, regardless of the exact value dmin_ξ. For this reason, we will generally consider such an interval to be "stable," and we will only examine transitions which involve iteration degrees strictly greater than k.

In principle, there are various kinds of discontinuities which can occur in the function (dmin_ξ), but we will restrict our attention to the following basic scenarios:

- an *increasing transition*, where $\mathrm{dmin}_\xi = j+1$ and $\mathrm{dmin}_\zeta = j$ for $\zeta \in (\xi - \epsilon, \xi)$;
- an *decreasing transition*, where $\mathrm{dmin}_\xi = j+1$ and $\mathrm{dmin}_\zeta = j$ for $\zeta \in (\xi, \xi - \epsilon)$.

As noted, these are not the only possible kinds of discontinuities; moreover, these two kinds of transitions can occur at the same point. And, more problematically, we might encounter essential discontinuities, for which every neighborhood $(\xi, \xi+\epsilon)$ (or $(\xi-\epsilon, \xi)$) contains ζ_1, ζ_2 such that $\mathrm{dmin}_{\zeta_1} \neq \mathrm{dmin}_{\zeta_2}$. However, it is not our intention to catalog and individually rule out every possible kind of degenerate behavior, and, for our purposes, it will suffice to analyze jump discontinuities of the kind described above.

For either of these transitions, note that value of dmin_ξ is greater than the value of dmin_ζ in the adjacent open interval. This is due to the fact that, the set

$$\{\xi : \mathrm{dmin}_\xi \leq j\}$$

is an open subset of $[0, 1)$. As such, increasing and decreasing transitions differ qualitatively in certain respects.

Increasing transitions can be detected by continuity from the left-hand limit $\mu_\xi = \lim_{\zeta \to \xi^-} \mu_\zeta$. Of course, in principle, the same holds for decreasing transitions and right-hand limits. However, the situation is not quite symmetric, since we will generally be attempting to solve for μ_ζ in an interval $\zeta \in (\xi, \xi+\epsilon)$ given the value of μ_ξ and not vice-versa. For this reason, decreasing transitions are slightly more complicated than increasing transitions.

Finally, before we proceed to discuss transitions at a more rigorous level, we point out once again that the asymptotic minimum degree depends on the solution to the process, and cannot be determined from the sample path. Hence, a change in the asymptotic minimum degree does not necessarily correspond to any specific event or hitting time.

Indeed, the true minimum degree $\mathbf{dmin}_t$ often changes from one iteration to the next, and hence a change in $\mathbf{dmin}_t$ does not indicate a transition point. On the other hand, changes on the asymptotic degree are "rare," and typically signal a qualitative change in behavior for the underlying process. However, since dmin_ξ is not observable from the sample path, then it is not generally possible to determine exactly (i.e. in which discrete time step) this change in behavior occurs.

Increasing Transitions

As noted above, increasing transitions can be detected by left-hand limits, and thus, from the point of view of computing a unique solution to the given system of differential equations, there is really nothing else we need to prove. That is, if we are able to compute a solution in some stable interval $\xi \in (\zeta_1, \zeta_2)$, for which $\mathrm{dmin}_\xi = j$, this solution extends by continuity to ζ_2. If $\mathrm{dmin}_{\zeta_2} > j$, then an increasing transition has "already occurred," and we may now attempt to extend the solution to an interval $[\zeta_2, \zeta_2 + \epsilon)$ using the values μ_{ζ_2} and dmin_{ζ_2}.

Hence, the stability conditions in §10.3.5 already provide necessary and sufficient conditions for an increasing transition to occur. Nevertheless, we shall supplement this analysis with the following simple proposition, which describes the possible values of $\mu_\xi(j)$ for which an increasing transition to $\mathrm{dmin}_\xi = j + 1 > k$ can occur.

Proposition 10.3.13. *Assume* $\mu_\xi(j) > \frac{1}{2k-j}$ *for* $j \geq k$ *and* $0 < \xi < 1$. *Then* $\mathrm{dmin}_\xi \leq j$.

Proof. Assume the contrary, so there exists $\xi \in (0, 1)$ such that $\mu_\xi(j) > \frac{1}{2k-j}$ and $\mathrm{dmin}_\xi = j + 1$. In this case $\mu_\xi(i) = 0$ for $i < j$, and thus by continuity we may choose $\epsilon > 0$ such that

$$\mu_\zeta(j) - \mu_\zeta(j-1) > \frac{1}{2k-j}$$

for $\zeta \in [\xi - \epsilon, \xi]$.

Now, recall that by proposition 10.3.3, we have

$$d\Lambda_\zeta(j) = (\mu_\zeta(j) - \mu_\zeta(j-1))d\sigma_\xi[R] - d\sigma_\zeta[L(j)].$$

We now claim that $d\Lambda_\zeta(j) > 0$ for $\zeta \in [\xi - \epsilon, \xi]$. First, since every iteration of any degree includes at least one random step, then $d\sigma_\zeta[R] > 0$, and hence we have the strict inequality

$$(\mu_\zeta(j) - \mu_\zeta(j-1))d\sigma_\zeta[R] > \frac{d\sigma_\zeta[R]}{2k-j}.$$

And, since since all but $o(m)$ iterations of degree j include $2k - j$ random steps, then we also have

$$d\sigma_\zeta[L(j)] \leq \frac{d\sigma_\zeta[R]}{2k-j}.$$

It follows that $d\Lambda_\zeta(j) > 0$ for $\zeta \in [\xi - \epsilon, \xi]$, which means that $\Lambda_\zeta(j)$ is strictly increasing in this interval. This in turn implies that $\Lambda_\xi(j) > 0$, which contradicts the assumption that $\text{dmin}_\xi = j + 1$, and the proof is complete. □

In particular, this proposition implies that in order for an increasing transition to $\text{dmin}_\xi = j + 1 > k$ to occur, we must have $\mu_\xi(j) \leq \frac{1}{2k-j}$. And, if this inequality is strict, meaning that $\mu_\xi(j) < \frac{1}{2k-j}$, then proposition 10.3.11 implies $\mu_{\xi^+} = j + 1$ as well, in which case we have a "clean" transition between two stable intervals of different degrees (i.e. a jump discontinuity).

Decreasing Transitions

Decreases in the asymptotic minimum degree cannot be detected by left-hand limits in this way, and the analysis of such transitions is therefore slightly more involved. However, based on propositions 10.3.11 and 10.3.14, we already have a certain amount of information about decreasing transitions.

Specifically, if $\text{dmin}_\xi = j + 1 > k$ and $\mu_\xi(j) < \frac{1}{2k-j}$ then proposition 10.3.11 ensures that $\text{dmin}_{\xi^+} = j + 1$ as well, so a decreasing transition cannot occur. Moreover, by 10.3.14,

$\mu_\xi(j) > \frac{1}{2k-j}$ cannot hold for any $\xi > 0$ such that $\mathrm{dmin}_\xi = j+1$. Hence, unless $\xi = 0$, the only possible state for which a decreasing transition can occur is if $\mu_\xi(j) = \frac{1}{2k-j}$.

Now, the fact that $\mu_\xi(j) = \frac{1}{2k-j}$ does not immediately imply a decreasing transition (although, in "typical" cases, a decreasing transition does indeed occur). In order to determine the behavior of dmin_{ξ^+} in this situation, we must take into account the values of $\mu_\xi(i)$ for $i > j$. The basic scenarios under which a decreasing transition takes place are listed in the following proposition.

Proposition 10.3.14. *Assume that* $\mathrm{dmin}_\xi = j+1 > k$. *Then:*

1. $\mathrm{dmin}_{\xi^+} \geq j$;
2. *if* $\mu_\xi(j) > \frac{1}{2k-j}$ *then* $\mathrm{dmin}_{\xi^+} = j$;
3. *if* $\mu_\xi(j) = \frac{1}{2k-j}$ *and*
$$\mu_\xi(j+1) > \frac{j(2k-j-1)}{(j+1)(2k-j)^2}, \tag{10.12}$$
then $\mathrm{dmin}_{\xi^+} = j$.

Proof. The first claim follows immediately from 10.3.11, since if $\mathrm{dmin}_\xi = j+1$ then $\mu_\xi(j-1) = 0$ and therefore $\mathrm{dmin}_{\xi^+} > j-1$. For the second claim, note that by proposition 10.3.14, the conditions $\mu_\xi(j) > \frac{1}{2k-j}$ and $\mathrm{dmin}_\xi = j+1 > k$ cannot occur for $\xi > 0$, and hence this is only possible in the initial state $\xi = 0$. In this case, by continuity, we have $\mu_\zeta(j) > \frac{1}{2k-j}$ for $\zeta \in (0, \epsilon)$, and hence $\mathrm{dmin}_\zeta \leq j$ for all such ζ, which implies that $\mathrm{dmin}_{\xi^+} \leq j$. And, by the first claim, this in turn implies that $\mathrm{dmin}_{\xi^+} = j$.

For the final claim, we shall prove that, under these assumptions, we must have $d\mu_\xi(j) > 0$, which implies that $\mu_\zeta(j) > \frac{1}{2k-j}$ in some neighborhood $\zeta \in (\xi, \xi+\epsilon)$. If this is indeed the case, we may then we may conclude by proposition 10.3.13 than $\mathrm{dmin}_\zeta \leq j$ in this interval, and therefore $\mathrm{dmin}_{\xi^+} = j$.

To establish that $d\mu_\xi(j) > 0$, we recall that, by proposition 10.3.9, if $\mathrm{dmin}_\xi = j$ and $\mu_\xi(j) = \frac{1}{2k-j}$, then we have $d\sigma_\xi[L(j+1)] = 0$ and
$$\theta_\xi = \frac{d\sigma_\xi[D]}{d\sigma_\xi[R]} = \frac{j}{2k-j}.$$

Using the basic equation for the $d\mu_\xi(j)$ in terms of random and deterministic selections from chapter 5, we may now compute

$$\begin{aligned} d\mu_\xi(j) &= \frac{((j+1)\mu_\xi(j+1) - j\mu_\xi(j))d\sigma_\xi[R] + \mu_\xi(j)d\sigma_\xi[D]}{1-\xi} \\ &= \left((j+1)\mu_\xi(j+1) - j\mu_\xi(j) + \left(\frac{j}{2k-j}\right)\mu_\xi(j)\right)\frac{d\sigma_\xi[R]}{1-\xi} \\ &= \left((j+1)\mu_\xi(j+1) - j\left(\frac{2k-j-1}{2k-j}\right)\mu_\xi(j)\right)\frac{d\sigma_\xi[R]}{1-\xi}. \end{aligned}$$

It follows that if $(j+1)\mu_\xi(j+1) > j\left(\frac{2k-j-1}{2k-j}\right)\mu_\xi(j)$, then $d\mu_\xi(j) > 0$. And, since $\mu_\xi(j) = \frac{1}{2k-j}$, this condition is equivalent to

$$\mu_\xi(j+1) > \frac{j(2k-j-1)}{(j+1)(2k-j)^2}.$$

□

Now, this proposition clearly does not cover all possible cases where $\mu_\xi(j) = \frac{1}{2k-j}$, since if $\mu_\xi(j+1) \leq \frac{j(2k-j-1)}{(j+1)(2k-j)^2}$, we are once again unable to predict the continuation dmin_{ξ^+}. Nevertheless, it is relatively straightforward to continue along these lines and determine dmin_{ξ^+} for a progressively larger set of specific situations. For instance, as one would expect, it can be shown that if the inequality in (10.12) is reversed, so $\mu_\xi(j+1) < \frac{j(2k-j-1)}{(j+1)(2k-j)^2}$, then $\mathrm{dmin}_{\xi^+} = j+1$ (although this fact is not quite immediate, and the proof requires more effort than for the above proposition). And, if strict equality holds, we must examine $\mu_\xi(j+2)$, and so on.

However, for the purposes of this chapter, proposition 10.3.14 will suffice. This is because in "typical" situations, if $\mu_\xi(j) = \frac{1}{2k-j}$ we will in fact have $\mu_\xi(j+1) > \frac{j(2k-j-1)}{(j+1)(2k-j)^2}$, and thus a decreasing transition will take place. Quite informally, this is because, in order for $\mu_\xi(j)$ to reach the transitional value $\frac{1}{2k-j}$ in the first place, $d\mu_\zeta(j)$ will (in "most" cases) have been strictly positive some neighborhood $\zeta \in (\xi - \epsilon, \xi)$. And, unless $d\mu_\zeta(j)$ drops to 0 at exactly the same time that $\mu_\xi(j)$ reaches $\frac{1}{2k-j}$, the inequality (10.12) will be satisfied.

10.3.7 Termination

In principle, by solving the system of differential equations derived above, we may compute a solution to μ_ξ for all $\xi \in [0,1)$ which corresponds to the sample path of the k-orientation algorithm on a random input path. However, ultimately, all we are interested is whether this algorithm terminates successfully, and this information cannot be ascertained from the values μ_ξ for $\xi \in [0,1)$.

On one hand, if the asymptotic minimum degree increases to $2k$ at some point $\xi < 1$, then we may conclude that the k-orientation heuristic terminates with failure.[9] Of course, this not imply that the input graph $\boldsymbol{\mathcal{G}}(\mu)$ is not k-orientable, since our heuristic is not guaranteed to succeed on all k-orientable graphs. But, it does imply that the k-orientability of $\boldsymbol{\mathcal{G}}(\mu)$ cannot be determined using this particular methodology.

On the other hand, if the solution to (μ_ξ) extends to all $\xi \in [0,1)$ without dmin_ξ ever reaching $2k$, this does not imply that the algorithm terminates successfully. All we can conclude is that w.e.h.p., for any fixed $\epsilon > 0$, the algorithm does not fail at any time $t < (1-\epsilon)m$.

In order to show that the algorithm actually terminates successfully, some additional work is required. In fact, the condition that $\mathrm{dmin}_\xi < 2k$ for all $\xi \in [0,1)$ is not sufficient to imply successful termination. In order to conclude that the algorithm terminates successfully, we require not only that $\mathrm{dmin}_\xi < 2k$, but also that dmin_ξ drops down to k and remains at most k for some fixed interval $\xi \in (1-\epsilon, 1)$.

Proposition 10.3.15. *Let μ be a distribution with first moment $M(\mu) < \infty$,*[10] *and assume*

[9] This is not quite immediate, since even if $\mathrm{dmin}_\xi = 2k$, the graph may be k-orientable if every remaining vertex has degree exactly $2k$. It is not difficult to show that this cannot occur for any $\xi < 1$; that is, if dmin_ξ reaches $2k$ at some $\xi < 1$ then we must have $\Lambda_\xi(2k) < 1$ (unless both $\xi = 0$ and $\Lambda_\xi(2k) = 1$ by specification), and therefore the average degree is strictly greater than $2k$, so the heuristic fails. We omit the proof of this fact, since, as mentioned above, the failure of this heuristic does not imply that $\boldsymbol{\mathcal{G}}(\mu)$ is not k-orientable, so we are only interested in conditions which imply that the heuristic succeeds.

[10] The assumption $M(\mu) < \infty$ may appear redundant, since any graph with average degree greater than $2k$ is no k-orientable. However, $M(\mu)$ is the first moment of the *residual* distribution, which corresponds to the second moment of the degree distribution by $M(\mu) = M_2(\lambda)/M(\lambda)$, and hence it is quite possible to have $M(\mu) = \infty$ even if the average true degree satisfies $M(\lambda) < 2k$.

that there exists a fixed $\epsilon > 0$ such that every w.e.h.p. solution to the k-orientability process satisfies $\mathrm{dmin}_\xi \leq k$ *for all* $\xi \in (1-\epsilon, 1)$.

Then any random graph $\mathbf{G} = \boldsymbol{\mathcal{G}}(\mu)$ *for which the maximum degree is* $O(m^{1/8-\Omega(1)})$, *and such that* $M(\mu_{\mathbf{G}}) \to M(\mu)$, *is k-orientable w.p.h.p.*

Proof. Note that in this case, we have $\theta_\xi = 1$ for $\xi \in (1-\xi, 1)$, and hence random and deterministic steps alternate as in the core algorithm. Moreover, the fact that $\mathrm{dmin}_\xi \leq k$ for $\xi \in (1-\epsilon, 1)$ implies that the residual distribution μ_ζ at time $\zeta = 1-\epsilon$ is such that the $(k+1)$-core of the random graph $\boldsymbol{\mathcal{G}}(\mu_\zeta)$ has size $o(m)$ w.e.h.p.

And, as shown in the previous chapter, the degree bound of $O(m^{1/8-\Omega(1)})$ implies that the k-core of this graph is w.p.h.p. empty, and hence the k-orientability process terminates successfully w.p.h.p. □

10.3.8 A Formal Theorem Statement

We conclude this chapter with a formal statement which describes the entire system of differential equations which governs the k-orientation process.

Theorem 10.3.16. *Consider the following system of differential equations.*

- *The state consists of:*
  1. *non-negative $\mathbb{R}$-valued functions $(\mu_\xi(i))$ for $i \in \mathbb{Z}^*$, $(\sigma_\xi[R])$, and $(\sigma_\xi[D])$, each of which have domain $\xi \in [0,1)$;*
  2. *the discrete function* (dmin_ξ) *defined by*
$$\mathrm{dmin}_\xi = \min\{i \geq 1 : \mu_\xi(i-1) > 0\};$$
  3. *the equations $\sigma_\xi[R] + \sigma_\xi[D] = \xi$ and $\sum_{i=0}^{\infty} \mu_\xi(i) = 1$ hold for all ξ.*
- *For $\xi > 0$, the differential equations are:*
  1. *if* $\mathrm{dmin}_\xi \leq k$ *then* $\frac{d\sigma_\xi[R]}{d\sigma_\xi[D]} = 1$;

2. *if* $\mathrm{dmin}_\xi = j + 1 > k$ *then*

$$\frac{d\sigma_\xi[D]}{d\sigma_\xi[R]} = \frac{j + 1 - 2k\mu_\xi(j)}{2k - j - 1} \tag{10.13}$$

3. *if* $i < \mathrm{dmin}_\xi - 1$ *then* $d\mu_\xi(i) = 0$*;*

4. *if* $i \geq \mathrm{dmin}_\xi$ *then*

$$d\mu_\xi(i) = \frac{((i+1)\mu_\xi(i+1) - i\mu_\xi(i))\, d\sigma_\xi[R] + \mu_\xi(i) d\sigma_\xi[D]}{1 - \xi}. \tag{10.14}$$

For a given distribution μ *with first moment* $M(\mu) < \infty$*, assume that every solution to the above system of differential equations, with initial condition* $\mu_0 = \mu$*, satisfies:*

1. $\mathrm{dmin}_\xi < 2k$ *for* $\xi \in [0, 1)$*, and*

2. $\mathrm{dmin}_\xi \leq k$ *for* $\xi \in (1 - \epsilon, 1)$*, for some fixed* $\epsilon > 0$*.*

Then the k-orientation heuristic succeeds w.p.h.p. for a random graph $\mathbf{G} = \boldsymbol{\mathcal{G}}(\mu)$ *satisfying* $M(\mu_{\mathbf{G}}) \to M(\mu)$*, and with maximum degree* $m^{1/8-\Omega(1)}$*.*

Proof. This is more or less immediate due to the various differential constraints derived throughout the section. In particular, the non-differential constraints hold by definition: $\sigma_\xi[R] + \sigma_\xi[D] = \xi$ is clear since every step is either deterministic or random, and $\sum_{i=0}^{\infty} \mu_\xi(i) = 1$ holds since the residual distribution is a proper probability distribution.

We may review the source of the four differential constraints listed above as follows. First, if $\mathrm{dmin}_\xi \leq k$, random and deterministic steps alternate, so we have $\theta_\xi = \frac{d\sigma_\xi[D]}{d\sigma_\xi[R]} = 1$. If $\mathrm{dmin}_\xi = j + 1 > k$, then (10.13) follows from proposition 10.3.10, provided the $\mu_\xi(j) \leq \frac{1}{2k-j}$. And, by proposition 10.3.13, for any $\xi > 0$ such that $\mathrm{dmin}_\xi = j + 1 > k$, we must indeed have $\mu_\xi(j) \leq \frac{1}{2k-j}$, hence this equation covers all possible cases.

The third constraint, that $d\mu_\xi(i) = 0$ for $i < \mathrm{dmin}_\xi - 1$ is not quite immediate. However, since $\mathrm{dmin}_{\xi^+} \geq \mathrm{dmin}_\xi - 1$ by proposition 10.3.14, we must at least have $\mu_\zeta(i) = 0$ for $\zeta \in (\xi, \xi + \epsilon)$ when $i < \mathrm{dmin}_\xi - 2$, which implies that $d\mu_\xi(i) = 0$. For the case $i = j - 1$, since true degree is one greater than residual degree, it suffices to show that $d\Lambda_\xi(j) = 0$.

In this case, we recall that, due to the fact that $\mu_\xi(j-1) = 0$, we have

$$d\Lambda_\xi(j) = \mu_\xi(j) d\sigma_\xi[R] - d\sigma_\xi[L(j)].$$

Again, if $\text{dmin}_\xi = j+1$ for $\xi > 0$, then $\mu_\xi(j) \leq \frac{1}{\min\{j, 2k-j\}}$, in which case proposition 10.3.9 implies that $d\sigma_\xi[L(j)] = \mu_\xi(j) d\sigma_\xi[R]$, and therefore $d\Lambda_\xi(j) = 0$.

Now, we note that these differential equations do not explicitly determine $d\mu_\xi(j)$ when $j+1 = \text{dmin}_\xi$. But, using the fact that the $\mu_\xi(i)$ must sum to 1, along with the constraint $d\mu_\xi(i) = 0$ for $i < j$, we may compute

$$d\mu_\xi(j) = d\xi - \sum_{i=j+1}^{\infty} d\mu_\xi(i).$$

Finally, by proposition 10.3.15, if the asymptotic minimum degree never reaches $2k$, and also $\text{dmin}_\xi \leq k$ for $\xi \in (1-\epsilon, 1)$, then proposition 10.3.15 implies successful termination.

□

10.4 Solving the System of Differential Equations

Theorem 10.3.16 constitutes a set of conditions which are sufficient (but not (necessarily) necessary) to establish that a random graph $\boldsymbol{\mathcal{G}}(\mu)$ is w.p.h.p. k-orientable. However, for any given distribution μ, in order to verify these conditions, we must actually solve these differential equations and then show that the asymptotic minimum degree satisfies $\text{dmin}_\xi \leq k$ for some interval $\xi \in (1-\epsilon, 1)$. Hence, while in principle we might consider the problem of determining the outcome of the k-orientation heuristic to be "solved," in practice, some additional work is required if we wish to actually invoke this theorem.

Moreover, the system of differential equations from theorem 10.3.16 is not particularly "easy" to solve. As usual, we have the issue of infinite-dimensionality, but based on the techniques developed in previous chapters, this issue is not overly problematic. A more substantial difficulty is the fact that the expression for $\theta_\xi = \frac{d\sigma_\xi[D]}{d\sigma_\xi[R]}$ is not continuous as a function of μ_ξ. In fact, although in the previous section, we were able to show that, in

"typical" situations, the continuation dmin_{ξ^+} is constant, and therefore these differential equations are stable, we have not covered all possible transitional scenarios. Hence, based on our current results, we cannot guarantee that this system of differential equations is uniquely solvable.

Due to these complications, we are not able to solve these differential equations in general form. In the next section, we will show that, for a certain restricted class of distributions, the fact that the k-orientation heuristic will succeed for the random graph $\boldsymbol{\mathcal{G}}(\mu)$ can be verified algebraically without explicitly solving any differential equations. However, for an arbitrary residual distribution μ, at present, the best technique we can offer for determining the outcome of this heuristic is to actually solve these differential equations on a case by case basis.

In this section, we discuss certain methods for computing such a solution. These methods make use of the Taylor approximations of the probability generating function which were introduced in the previous chapter for the purpose of studying the k-core of $\boldsymbol{\mathcal{G}}(\mu)$. As we shall see, despite the fact that the system of differential equations from theorem 10.3.16 is both infinite-dimensional and discontinuous, using these techniques, finding a solution is not intractably difficult (in most cases).

Nevertheless, as noted above, we are not able to find a "clean" general solution for an arbitrary distribution μ, and hence we will not attempt to produce a rigorous theorem statement which determines whether the k-orientation will succeed based on algebraic properties of μ. Indeed, sufficient conditions for the success of this heuristic have already been derived in theorem 10.3.16. In this section, our objective is merely to illustrate, primarily by example, how generating functions can be used to verify the conditions of theorem 10.3.16 for a given residual distribution μ.

10.4.1 Overall Strategy

At a high level, the execution of the k-orientation can be characterized in terms of stable intervals, during which the asymptotic minimum degree remains constant, and

transition points, which are discontinuities in the discrete function (dmin_ξ). Moreover, during any stable interval for which $\mathrm{dmin}_\xi = j+1$, the ratio $\theta_\xi = \frac{d\sigma_\xi[D]}{d\sigma_\xi[R]}$ is Lipschitz continuous as a function of $\mu_\xi(j)$. Accordingly, our strategy for solving this system of differential equations is to compute a solution piecewise, one stable interval at a time.

For a number of reasons, our current set of rigorous results cannot guarantee that this piecewise solution strategy will succeed for every possible choice of residual distribution μ. These issues are briefly touched upon below; for now, we shall give an overview of how this strategy will work in "well-behaved" situations.

At the outset, we are faced which what appears to be a "Catch-22" situation: in order to show that an interval $(0, \zeta)$ is stable, we must compute μ_ξ in for $\xi \in (0, \zeta)$, but we cannot perform this computation without first assuming that this interval is indeed stable. We may resolve this problem as follows. The stability and transitional conditions from §10.3.5 and §10.3.6 will (in most cases) allow us to determine the continuation dmin_{0^+} based on the initial state μ_0. And, if are able to show that dmin_{0^+} is equal to a given constant, then, by definition, there must exist some stable interval $\xi \in (0, \epsilon)$.

Of course, we cannot yet determine how far this stable interval extends. But, we do know that its length is strictly positive. Therefore, we will temporarily ignore the question of how long dmin_ξ remains constant. Instead, we will simply compute a preliminary solution to μ_ξ for all $\xi \in (0, 1)$ under the assumption the entire interval $\xi \in (0, 1)$ is stable.

Now, in any stable interval for which $\mathrm{dmin}_\xi = j+1$, we have $d\sigma_\xi[L(i)] = 0$ for $i > j+1$, which implies that deterministic selections have degree at most $j+1$ (i.e. residual degree at most j). In this case, we may compute the values of $\mu_\xi(i)$ for $i > j$ using the methods from the previous chapter (specifically corollary 9.2.7) provided that we can determine the (weighted) total number of deterministic and random selections. Moreover, the fact that $\mathrm{dmin}_\xi = j+1$ also implies that $\mu_\xi(i) = 0$ for $i < j$, and therefore we may compute the only remaining coordinate of the residual distribution by

$$\mu_\xi(j) = 1 - \sum_{i=j+1}^{\infty} \mu_\xi(i).$$

Hence, in any stable interval, corollary 9.2.7 will yield a formula for $\mu_\xi(j)$ based on random and deterministic selections. And, by equation (10.9) from proposition 10.3.10 we also have a formula for the ratio $\theta_\xi = \frac{d\sigma_\xi[D]}{d\sigma_\xi[R]}$ as a Lipschitz continuous function of $\mu_\xi(j)$. Combining these two formulas thus results in a finite-dimensional system of differential equations which can be solved routinely.

This will yield a preliminary solution to μ_ξ for all $\xi \in (0, 1)$, under the assumption that $\text{dmin}_\xi = j + 1$ for this entire interval. Now, this assumption will typically not hold, but, this preliminary solution will allow us to detect the first transition point ζ. Indeed, proposition 10.3.11 implies that a transition cannot occur while $0 < \mu_\xi(j) < \frac{1}{2k-j}$, and our potential solution will extend at least until the smallest ζ such that either $\mu_\zeta(j) = 0$ or $\mu_\zeta = \frac{1}{2k-j}$.

At this point, we can repeat this entire procedure to compute a solution in a second stable interval $\xi \in (\zeta, \zeta_1)$ based on the transitional state μ_ζ, and so on.

Potential Difficulties

As noted above, based on our current set of rigorous results, we cannot guarantee that this strategy will succeed for any choice of residual distribution μ. Of course, there are various practical concerns which might prevent us from actually carrying out the computations outlined above. However, there are also certain theoretical issues which we have left unresolved and which, in principle, might imply that this piecewise solution strategy can fail in theory as well as practice.

First, since proposition 10.3.14 does not cover all possible transitional scenarios, we might encounter a state where the continuation dmin_{ξ^+} cannot be determined from the residual distribution μ_ξ. Now, in §10.3.6, we claim to have covered the "majority" of transitional scenarios, and it appears straightforward to generalize proposition 10.3.13 to determine dmin_{ξ^+} in cases which we have not covered. Nevertheless, these are both non-rigorous claims, and therefore we cannot, at this point, rule out the possibility that the solution will enter a state μ_ξ for which dmin_{ξ^+} cannot be determined, even in theory, from

the residual distribution.

Second, the piecewise solution strategy will evidently only succeed if there are finitely many transition points, and, based on our current results, we cannot guarantee that this must be the case. Again, non-rigorous arguments suggest that an infinite number of transition points is not a possibility. Briefly, this is because if the number of transition points is infinite, then they must accumulate somewhere in $[0, 1]$. This accumulation point will in turn be an essential discontinuity in the discrete function (dmin_ξ). Now, if it is indeed the case that proposition 10.3.13 can be generalized to show that the continuation dmin_{ξ^+} must be equal to a fixed constant for any state μ_ξ, this would rule out the possibility of essential right-discontinuities. And, it "appears" that essential left-discontinuities can be ruled out similarly. But, this argument has not been made rigorous (either in this chapter or on "scratch paper"), and therefore we cannot formally discount the possibility of encountering an essential discontinuity.

In any case, even if our intuition is correct and neither of these theoretical problems turn out to be fatal, in practice, the strategy outlined above might nevertheless be infeasible for certain choices of μ. Indeed, as a practical matter, if a given solution exhibits, say, 10^{10} transition points, then the fact that there exists a unique solution with finitely many transition points is largely irrelevant. Either way, we will not be able to compute a piecewise solution, and if we wish to determine the outcome of the k-orientation heuristic, we are going to have to use a different technique.

As such, we will not concern ourselves with the possibility of degenerate behavior, and instead we will simply show how to compute a solution to this system of differential equations under "favorable" conditions. Indeed, we are ultimately only interested in whether the asymptotic minimum degree remains bounded by $\text{dmin}_\xi \leq k$ for some interval $\xi \in (1-\epsilon, 1)$, and there are obviously other possible strategies for attempting to establish this fact without explicitly computing μ_ξ for all $\xi \in [0, 1)$. Hence, if we encounter a particularly difficult case, our efforts are better spent trying to develop a different method for predicting this terminal behavior, rather than proving that it is theoretically possible to compute a solution through

brute force.

10.4.2 Solution in a Stable Interval

We now explain how the generating functions technique from the previous chapter can be used to compute a solution to the system of differential equations governing the k-orientation process for a single stable interval. For simplicity, we will consider a stable interval $\xi \in (0, \zeta)$, but we note that this is without loss of generality, since the same technique can be used for any stable interval $\xi \in (\zeta_1, \zeta_2)$ by changing variables appropriately

We begin by reviewing some basic definitions and results from the previous chapter. First, for any distribution μ, the probability generating function is given by

$$\psi_\mu(z) = \sum_{i=0}^{\infty} z^i \mu_\xi(i).$$

We also make use of two functional transformations of an arbitrary (real-analytic) function $\psi(z)$. First, for any j, we denote the j'th Taylor approximation of $\psi(1)$ about z by

$$(\gamma_j \psi)(z) = \psi(z) + \sum_{i=1}^{j} \frac{(1-z)^i \psi^{(i)}(z)}{i!},$$

where $\psi^{(i)}(z)$ denotes the i'the derivative. Also, for any $(x, y) \in \mathbb{R}^2$, we define

$$(\alpha_{(x,y)} \psi)(z) = \frac{\psi(x + z - x \cdot z) - y}{1 - y}.$$

Typically we will have both $x, y < 1$, in which case this transformation corresponds to "moving the origin" from the point $(0, 1)$ to (x, y) and rescaling the unit square as described in chapter 6.

Using corollary 9.2.7 of the previous chapter, for any removal process satisfying $d\sigma_\xi[D(i)]$ for $i > j + 1$, these tools can be used to determine the residual distribution μ_ξ as follows. First, recall that the total random and deterministic weight, as defined in chapter 5, is given by $w_\xi[R] = \int_{\zeta=0} \frac{d\sigma_\xi[R]}{1-\xi}$ and similarly for $w_\xi[D]$. We then change variables by defining

$$x_\xi = 1 - e^{-w_\xi[R]} \quad \text{and} \quad y_\xi = 1 - e^{-w_\xi[D]}.$$

Provided that $d\sigma_\xi[D(i)] = 0$ for $i > j+1$ during an interval $\xi \in (0,\zeta)$, then every w.e.h.p. solution to μ_ξ will satisfy

$$(\gamma_j \psi_{\mu_\xi})(z) = (\alpha_{(x_\xi, y_\xi)} \gamma_j \psi_\mu)(z). \tag{10.15}$$

This function uniquely determines the values $\mu_\xi(i)$ for all $i > j$, as well as the total

$$\sum_{i=0}^{j} \mu_\xi(i) = (\gamma_j \psi_{\mu_\xi}(0)).$$

For the k-orientability process, if $\mathrm{dmin}_\xi = j+1 > k$ for $\xi \in (0,\zeta)$, then we the condition $d\sigma_\xi[D(i)] = 0$ for $i > j+1$ will hold in this interval, and hence this formula is applicable. Moreover, $\mathrm{dmin}_\xi = j+1$ also implies that $\mu_\xi(i) = 0$ for $i < j$, and hence this formula yields

$$\mu_\xi(j) = (\alpha_{(x_\xi, y_\xi)} \gamma_j \psi_\mu)(0) = \frac{(\gamma_j \psi_\mu)(x_\xi) - y_\xi}{1 - y_\xi}. \tag{10.16}$$

Therefore, provided that we can compute x_ξ and y_ξ, every coordinate of the residual distribution can be determined precisely using generating functions in this way. Since x_ξ and y_ξ depend on the ratio θ_ξ of deterministic to random steps, a complete solution in a stable interval can be computed as follows.

Proposition 10.4.1. *For any distribution μ, and any $j \geq k$ let us denote the j'th variant of the p.g.f. ψ_μ defined above by $\varphi_j(z) = (\gamma_j \psi_\mu)(z)$.*

And, let $y(x)$ denote the (unique) solution to the differential equation

$$y'(x) = \frac{y(x) - \frac{2k\varphi_j(x) - j - 1}{2k - j - 1}}{1 - x} \tag{10.17}$$

for $x \in [0,1)$, with initial condition $y(0) = 0$.

If the distribution μ is such that the k-orientation process satisfies $\mathrm{dmin}_{0^+} = j+1$ when the initial state is $\mu_0 = 0$, then:

1. *there exists a stable interval $\xi \in (0,\zeta)$ for which that $\mathrm{dmin}_\xi = j+1$;*
2. *the equation $y_\xi = y(x_\xi)$ holds for $\xi \in [0,\zeta)$;*

3. the termination point of this stable interval satisfies either $y(x_\zeta) = \varphi_j(x_\zeta)$ *or*

$$y(x_\zeta) = \frac{\varphi_j(x_\zeta) - \frac{1}{2k-j}}{1 - \frac{1}{2k-j}}. \tag{10.18}$$

Proof. First, the fact that $\mathrm{dmin}_{0^+} = j+1$ implies that there exists some such stable interval $(0, \epsilon)$, and the only question is how long this stable interval lasts. Moreover, during this stable interval, $\mu_\xi(j)$ is determined by (10.16) as discussed above. Hence, the ratio of deterministic to random steps is

$$\theta_\xi = \frac{j+1-2k\mu_\xi(j)}{2k-j+1} = \frac{j+1-2k\left(\frac{\varphi_j(x_\xi)-y_\xi}{1-y_\xi}\right)}{2k-j+1}.$$

Since $x_\xi = 1 - e^{-w_\xi[R]}$, then we may compute

$$dx_\xi = e^{-w_\xi[R]} dw_\xi[R] = (1-x_\xi) dw_\xi[R],$$

and similarly for dy_ξ, so we have

$$\begin{aligned}\frac{dy_\xi}{dx_\xi} = \frac{(1-y_\xi)dw_\xi[D]}{(1-x_\xi)dw_\xi[R]} = \left(\frac{1-y_\xi}{1-x_\xi}\right)\theta_\xi &= \left(\frac{1-y_\xi}{1-x_\xi}\right)\left(\frac{j+1-2k\left(\frac{\varphi_j(x_\xi)-y_\xi}{1-y_\xi}\right)}{2k-j+1}\right) \\ &= \frac{(j+1)(1-y_\xi) - 2k(\varphi_j(x_\xi)-y_\xi)}{(2k-j+1)(1-x_\xi)} \\ &= \frac{y_\xi - \frac{2k\varphi_j(x_\xi)-j-1}{2k-j-1}}{1-x_\xi}.\end{aligned}$$

We may therefore express y_ξ as a function of x_ξ by $y : x_\xi \mapsto y_\xi$, and since $y_0 = x_0 = 0$, this function must satisfy (10.17) with initial condition $y(0) = 0$.

Finally, by proposition 10.3.11, this stable interval must extend until either $\mu_\zeta(j) = 0$ or $\mu_\zeta(j) = \frac{1}{2k-j}$. Using (10.16), we easily verify that the case $\mu_\zeta(j) = 0$ corresponds to $y(x_\zeta) = \varphi_j(x_\zeta)$, and the second case corresponds to $\frac{\varphi_j(x_\zeta)-y_\zeta}{1-y_\zeta} = \frac{1}{2k-j}$, which is equivalent to

$$y(x_\zeta) = \frac{\varphi_j(x_\zeta) - \frac{1}{2k-j}}{1 - \frac{1}{2k-j}}.$$

□

Transitions

When we reach the transition point ζ, by changing variables appropriately, we may use the same technique to compute μ_ξ in the next stable interval (ζ, ζ_1). The change of variables is fairly routine, so we will not go into too much detail.

However, we note that if ζ is an increasing transition, and the asymptotic minimum degree increases to $\mathrm{dmin}_\zeta = j + 2$, then the condition $d\sigma_\xi[D(i)] = 0$ for $i > j + 2$ will hold in first stable interval as well as the second. In this case, the distribution μ_ζ will satisfy

$$(\gamma_{j+1}\psi_{\mu_\zeta})(z) = (\alpha_{(x_\xi, y_\xi)}\gamma_{j+1}\psi_\mu)(z),$$

and thus we may use the function $\varphi_{j+1}(z) = (\gamma_{j+1}\psi_\mu)(z)$ to compute the solution for this stable interval, using the same differential equations as proposition 10.4.1 (and, of course, substituting $\mathrm{dmin}_\xi = j + 2$ for $\mathrm{dmin}_\xi = j + 1$ in these equations).

If ζ is a decreasing transition, the function φ_{j-1} cannot be computed in the same way. But, in this case, we have $\mu_\zeta(j-1) = 0$, and hence

$$(\gamma_{j-1}\psi_{\mu_\zeta})(0) = 0.$$

It follows that the function φ_{j-1} must satisfy $(\alpha_{(x_\zeta, y_\zeta)}\varphi_{j-1})(0) = 0$, and therefore

$$\varphi_{j-1}(x_\zeta) = y_\zeta.$$

We may thus reconstruct the function $\varphi_{j-1}(z)$ from $\varphi_j(z)$ in this interval using the differential equation

$$\varphi_j(z) = \varphi_{j-1}(z) + \frac{(1-z)\varphi'_{j-1}(z)}{j}$$

from corollary 9.2.7 from the previous chapter. Hence, even for a decreasing transition, it is not necessary to start "from scratch" for this second stable interval.

10.4.3 An Example

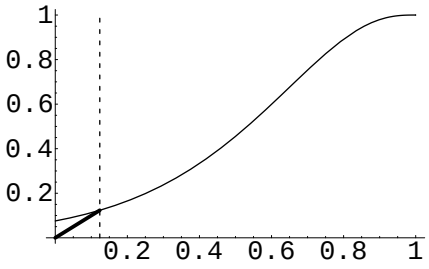
We now illustrate how proposition 10.4.1 can be used to compute a complete solution to the k-orientation process. For this illustrative example, we will consider the case $k = 3$,

and we will use the Poisson distribution with mean 5.72 as our residual distribution. This graph $\boldsymbol{\mathcal{G}}(\mu)$ in this case corresponds to the the Erdős-Rényi random graph $\mathbf{G}_{n,p}$ with $p = 5.72$. Also, the 4-core of such a graph has average degree 5.963 (as we may compute using the results of the previous chapter), which is somewhat close to maximum possible average degree of 6 for any 3-orientable graph.

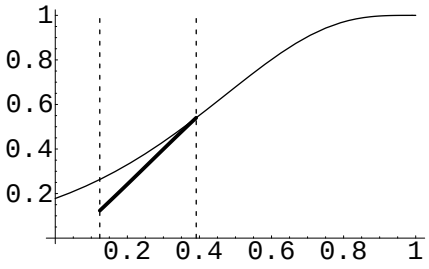
As for the k-core algorithm (as well as the giant component from chapter 6), the solution is best understood visually by actually plotting the relevant generating functions, along with the trajectory of the point (x_ξ, y_ξ) which represents the total weight of random and deterministic steps as the algorithm executes.

Now, if $\text{dmin}_\xi \leq k = 3$, then the ratio θ_ξ of deterministic and random steps is equal to 1, so it is not necessary to take the exact value of dmin_ξ into account. And, if dmin_ξ reaches $2k = 6$, then the algorithm fails. Hence, there are only three situations to consider: $\text{dmin}_\xi \leq 3$, $\text{dmin}_\xi = 4$, and $\text{dmin}_\xi = 5$. For these three cases, the relevant p.g.f. variants are φ_2, φ_3, and φ_4, respectively.

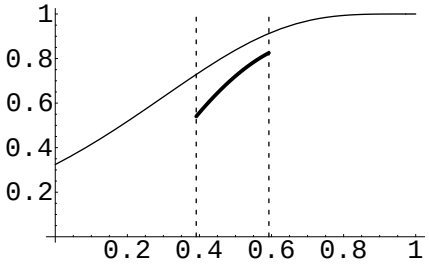
The algorithm begins with $\text{dmin}_\xi \leq 3$, and in this case we have $\theta_\xi = 1$, since random and deterministic steps alternate as in the k-core algorithm. This yields the simple differential equation $y'(x) = \frac{1-y(x)}{1-x}$, and since initially $y_0 = x_0 = 0$, then the point (x_ξ, y_ξ) moves in a straight line of slope 1 from $(0, 0)$ to $(1, 1)$. This continues until the trajectory intersects the curve $\varphi_2(x)$, at which point dmin_ξ reaches 4, and we have an an increasing transition.



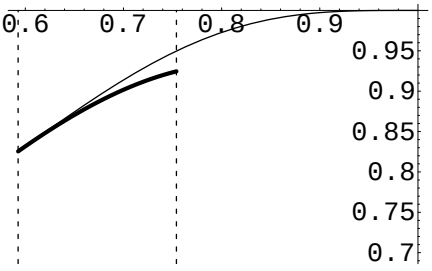
At this point, the ratio of deterministic to random selections is no longer 1. Instead, we must use equation (10.17) to compute this ratio. With $j+1 = 4$ and $k = 3$, we thus have $y'(x) = \frac{y(x)-3\varphi_3(x)+2}{1-x}$. This trajectory continues until $\mu_\xi(3) = \frac{\varphi_3(x_\xi)-y_\xi}{1-y_\xi}$, reaches either 0 or $\frac{1}{2k-j} = 1/3$. In this particular case the first condition occurs before the second, and a transition to $\text{dmin}_\xi = 5$ occurs at the point where the trajectory intersects the curve φ_3.



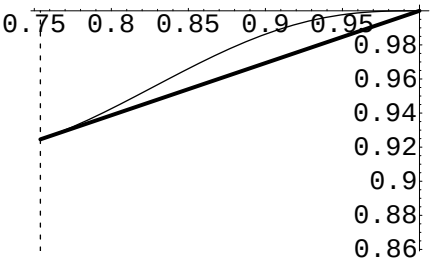
For $\mathrm{dmin}_\xi = 5$, the analysis is the same, except that we now use the function φ_4 and the trajectory is given by $y'(x) = \frac{y(x)-6\varphi_4(x)+5}{1-x}$. Note that, if dmin_ξ reaches 6, since in this case this case $6 = 2k$. Hence, if the trajectory intersects the curve φ_4, we have a failure. For the parameters we have chosen, this does not occur; instead, the next transition occurs at the point where $\mu_\xi(4) = \frac{\varphi_4(x_\xi)-y_\xi}{1-y_\xi} = \frac{1}{2\cdot 3-4} = 1/2$, and the dmin_ξ drops back down to 4.



A new complication now arises: in order to track the algorithm with $\mathrm{dmin}_\xi = 4$, we need the curve φ_3. This curve cannot be computed directly from the initial residual distribution, since endpoints of residual degree 4 have been subject to deterministic selection while dmin_ξ was equal to 5. However, as described above, we can reconstruct this function algebraically using φ_4, combined with the fact that $\mu_\xi(3) = 0$, and therefore $\varphi_3(x_\xi) = y_\xi$, at the decreasing transition point. We then proceed with $\mathrm{dmin}_\xi = 4$ until the next decreasing transition.



We are now back to $\mathrm{dmin}_\xi = 3 \leq k$, and therefore $\theta_\xi = 1$, so the trajectory once again moves in a straight line towards the point $(1, 1)$. We thus reconstruct the function φ_3 from φ_4 and plot this straight line, which extends all the way until termination.



Since $\mathrm{dmin}_\xi \leq k$ holds for all ξ in the last stable interval, then proposition 10.3.15 implies that the algorithm terminates successfully w.p.h.p. The entire execution of the k-orientation heuristic in this example, along with the relevant generating functions, is shown in figure 10.2.

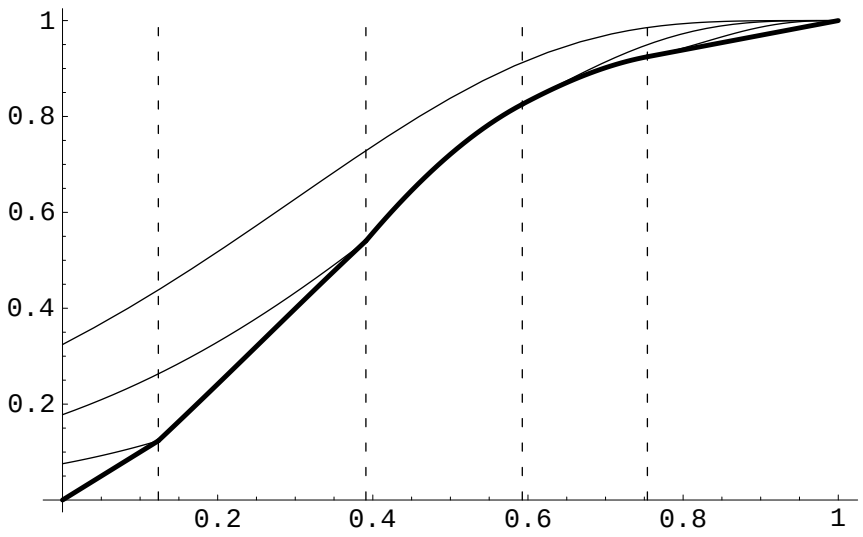


Figure 10.2: The execution of the k-orientation heuristic for $k = 3$, where μ is the Poisson distribution with mean 5.72. In this case, there are 4 transition points, and the algorithm terminates successfully.

10.5 k-Orientability for a Specific Family of Distributions

The generating functions technique described above greatly facilitates the task of explicitly solving the system of differential equations in theorem 10.3.16. Nevertheless, we still do not have a completely satisfactory solution to the k-orientability problem. Indeed, theorem 10.3.16 can generally only provide one-sided results, since the failure of this particular heuristic does not imply that $\boldsymbol{\mathcal{G}}(\mu)$ is not k-orientable. Moreover, the ability to solve these differential equations on a case by case basis does not provide any high-level insight into what general characteristics of a residual distribution μ determine whether the corresponding random $\boldsymbol{\mathcal{G}}(\mu)$ is k-orientable.

In this section we shall show that, if certain regularity conditions are imposed on μ, these two issues can be overcome. For this restricted class of distributions, we will demonstrate that the k-orientation heuristic will succeed provided that the average degree of the $(k+1)$-core of $\boldsymbol{\mathcal{G}}(\mu)$ is strictly less than $2k$. Now, if the $(k+1)$-core of $\boldsymbol{\mathcal{G}}(\mu)$ is has average degree strictly greater than $2k$, then we may immediately conclude of $\boldsymbol{\mathcal{G}}(\mu)$

is not k-orientable. Hence, for this particular class of distributions, the k-orientability property exhibits a sharp threshold. Moreover, using the techniques of the previous chapter, the average degree of the $(k+1)$-core can be computed fairly easily without solving any differential equations, and thus we have a relatively concise algebraic characterization of this k-orientability threshold.

The argument used to achieve this result is based on the behavior of the average degree. When the excess degree reduction heuristic was introduced in section 10.2, as informal justification, we observed that iterations of degree greater than k typically result in a ratio of $2k$ endpoints per vertex removed. And, if the average degree is strictly less than $2k$, this will cause the average degree to decrease each iteration.

In fact, due to proposition 10.3.1, the ratio of $2k$ endpoints per vertex holds for all but $o(m)$ iterations of degree $i \geq k$. Moreover, if the asymptotic minimum degree satisfies $\text{dmin}_\xi > k$, we have $d\sigma_\xi[L(i)] = 0$ for $i \leq k$, which in turn implies that, as long as dmin_ξ remains strictly greater than k, all but $o(m)$ iterations will have degree k or greater, and thus the average degree will decrease.

In particular, if the average degree is strictly less than $2k$ when dmin_ξ first reaches k (i.e. when the $(k+1)$-core is found), then the average degree will decrease while $k < \text{dmin}_\xi < 2k$. Now, obviously, if the average degree is decreasing, then it will remain below $2k$, and hence the minimum degree must remain below $2k$ as well. Moreover, the ratio $2k$ endpoints per vertex can not go on indefinitely, since eventually we will run out of endpoints. Thus, in this case, dmin_ξ must at some point drop back down to k.

When this occurs, it is no longer necessarily the case that the average degree decreases each iteration. However, as we shall see, by placing relatively simple restrictions on the residual distribution μ, we can guarantee that, once a decreasing transition of the form $\text{dmin}_\xi = k+1$ and $\text{dmin}_{\xi^+} = k$ occurs, the asymptotic minimum degree will satisfy $\text{dmin}_\zeta \leq k$ for all $\zeta \in (\xi, 1)$. And, by proposition 10.3.15, this implies that the k-orientation heuristic terminates successfully w.p.h.p.

For any distribution μ, and any i, let us now define

$$\mu^*(i) = \frac{\mu(i)}{\sum_{j \geq i} \mu(j)} \tag{10.19}$$

whenever the sum in the denominator is positive. Intuitively, for a residual distribution μ, the value $\mu^*(i)$ represents the number of endpoints of residual degree i, as a fraction of the total number of endpoints of residual distribution at least i. In terms of the values $\mu^*(i)$, the conditions which guarantee the behavior described above are the following.

Theorem 10.5.1. *Let μ be a distribution with $M(\mu) < \infty$, and such that:*

1. *the $(k+1)$-core of $\boldsymbol{\mathcal{G}}(\mu)$ w.e.h.p. has average degree less than $2k$;*
2. *$\mu^*(i+1) \geq \mu^*(i)$ for all $i > k$.*

Then any random graph $\mathbf{G} = \boldsymbol{\mathcal{G}}(\mu)$ with maximum degree $O(m^{1/8-\Omega(1)})$, and satisfying $M(\mu_{\mathbf{G}}) \to M(\mu)$, is w.p.h.p. k-orientable.

The proof of this theorem is organized as follows. First, we some preliminary lemmas related to the condition $\mu^*(i+1) \geq \mu^*(i)$ in §10.5.1 and §10.5.2. Then, in §10.5.3, we make rigorous the average degree argument discussed above. We then combine these lemmas in §10.5.4

10.5.1 The Condition $\mu^*(i+1) \geq \mu^*(i)$

We begin with the following lemma, which lists two key consequences of the condition $\mu^*(i+1) \geq \mu^*(i)$.

Lemma 10.5.2. *If $\mu^*(i+1) \geq \mu^*(i)$ holds initially then:*

1. *$\mu^*_\xi(i+1) \geq \mu^*_\xi(i)$ holds for all i and all $\xi \in [0,1)$;*
2. *for any i and $\xi \in [0,1)$ such that $\mathrm{dmin}_\xi < i$ and $\mu^*_\xi(i) < 1$, we have $d\mu^*_\xi(i) > 0$.*

Proof. We first compute the change $\frac{\partial \mu^*(i)}{\partial w_\xi[R]}$ in $\mu^*_\xi(i)$ during random selections. For this purpose, let us define $\rho_\xi(i) = \sum_{j \geq i} \mu_\xi(i)$, and as usual we have $\frac{\partial \mu(i)}{\partial w_\xi[R]} = (i+1)\mu_\xi(i+1) - i\mu_\xi(i)$, and hence

$$\frac{\partial \rho_\xi(i)}{\partial w_\xi[R]} = \sum_{j \geq i} \frac{\partial}{\partial w_\xi[R]} \mu_\xi(i) = -i\mu_\xi(i).$$

Since $\mu^*_\xi(i) = \frac{\mu_\xi(i)}{\rho_\xi(i)}$, we may compute

$$\begin{aligned}\frac{\partial \mu^*_\xi(i)}{\partial w_\xi[R]} &= \frac{(i+1)\mu_\xi(i+1) - i\mu_\xi(i)}{\rho_\xi(i)} - \mu^*_\xi(i)\left(\frac{-i\mu_\xi(i)}{\rho_\xi(i)}\right) \\ &= (i+1)\mu^*_\xi(i+1)\left(\frac{\rho_\xi(i+1)}{\rho_\xi(i)}\right) - i\mu^*_\xi(i) + i(\mu^*_\xi(i))^2,\end{aligned}$$

and, since $\rho_\xi(i+1) = \rho_\xi(i) - \mu_\xi(i)$, then this yields

$$\frac{\partial \mu^*(i)}{\partial w_\xi[R]} = (1 - \mu^*_\xi(i))\left((i+1)\mu^*_t(i+1) - i\mu^*_t(i)\right). \tag{10.20}$$

Now, if $\mu^*_\xi(i+1) \geq \mu^*_\xi(i)$, then this quantity is strictly positive unless $\mu^*_\xi(i) = 1$. And, since endpoints of residual degree i are subject only to random selection whenever $\mathrm{dmin}_\xi \leq i$, then in this case we have $d\mu^*_\xi(i) > 0$. Hence, the second claim follows from the first.

To prove the first claim, we consider three cases. First, if $\mathrm{dmin}_\xi \leq i$, then both $\mu^*_\xi(i+1)$ and $\mu^*_\xi(i)$ only change during random selections. Second if $\mathrm{dmin}_\xi = i+1$, then $\mu^*_\xi(i)$ only changes during random selections; but, in this situation, deterministic selections of degree $i+1$ (i.e. residual degree i) can only decrease $\mu^*_\xi(i)$, and therefore increase $\mu^*_\xi(i+1) - \mu^*_\xi(i)$. And in the third case, if $\mathrm{dmin}_\xi \geq i+1$, then $\mu^*_\xi(i) = \mu_\xi(i) = 0$, and hence clearly $\mu^*_\xi(i+1) \geq \mu^*_\xi(i)$.

It therefore suffices to prove that

$$\frac{\partial}{\partial w_\xi[R]} \mu^*_\xi(i+1) - \mu^*_\xi(i) \geq 0.$$

To simplify this computation, let us define $r_0 = \mu^*_\xi(i)$, $r_1 = \mu^*_\xi(i+1)$, and $r_2 = \mu^*_\xi(i+2)$.

We now compute

$$\begin{aligned}\frac{\partial}{\partial w_\xi[R]}\left(\mu^*_\xi(i+1)-\mu^*_\xi(i)\right) &= (1-r_1)((i+2)r_2-(i+1)r_1)-(1-r_0)((i+1)r_1-ir_0)\\ &= \Big((i+2)r_2+ir_0\Big)-r_1r_2-(i+1)r_1(r_2-r_1)+r_1r_0-ir_0(r_1-r_0)\\ &= \Big((i+2)r_2+ir_0\Big)-r_1\Big((i+1)(r_2-r_1)+(r_2-r_0)\Big)-ir_0(r_1-r_0)\\ &\geq \Big((i+2)r_2+ir_0\Big)-(i+2)r_1-ir_0\\ &\geq 0.\end{aligned}$$

□

10.5.2 Decreasing Transitions

We now demonstrate that the condition $\mu^*(i+1) \geq \mu^*(i)$ implies that, for any $j+1 \geq k$, once a decreasing transition from $\mathrm{dmin}_\xi = j+1$ to $\mathrm{dmin}_{\xi^+} = j$ occurs, the asymptotic minimum degree cannot increase back up to $j+1$.

Lemma 10.5.3. *If $\mu^*(i+1) \geq \mu^*(i)$ holds initially then, for any $j > k$, there exists at most one $\xi \in [0,1)$ such that $\mathrm{dmin}_\xi = j+1$ and $\mathrm{dmin}_{\xi^+} = j$, and in this case, we have $\mathrm{dmin}_\zeta < j$ for all $\zeta \in (\xi, 1)$.*

Proof. By proposition 10.3.13 any $\xi \in (0,1)$ with $\mu_\xi(j) > \frac{1}{2k-j}$, must satisfy $\mathrm{dmin}_\xi \leq j$. And, if $\mathrm{dmin}_\xi = j+1$, then we have $\mu_\xi(i) = 0$ for $i < j$ and in this case $\mu^*_\xi(j) = \mu_\xi(j)$. It follows that $\mathrm{dmin}_\xi \leq j$ holds whenever $\mu^*_\xi(j) > \frac{1}{2k-j}$ as well.

Moreover, as shown above, $d\mu^*_\xi(j) > 0$ whenever $\mathrm{dmin}_\xi \leq j$, and it follows that μ^*_ξ is strictly increasing for all values of $\xi \in (0,1)$ for which $\mu^*_\xi(j) > \frac{1}{2k-j}$. Hence, once $\mu^*_\xi(j)$ exceeds $\frac{1}{2k-j}$, we have $\mu^*_\zeta(j) > \frac{1}{2k-j}$ and therefore $\mathrm{dmin}_\zeta \leq j$ for all $\zeta \in (\xi, 1)$.

Finally, by proposition 10.3.11, if $\mathrm{dmin}_\xi = j+1$ and $\mathrm{dmin}_{\xi^+} = j$, then we must have $\mu_\xi(j) \geq \frac{1}{2k-j}$. In this situation, since $\mathrm{dmin}_{\xi^+} \leq j$ then we have $d\mu^*_\zeta(j) > 0$ for ζ in some neighborhood $(\xi, \xi+\epsilon)$, and it follows that $\mu^*_\zeta(j) > \frac{1}{2k-j}$ in this neighborhood. Hence, as shown above $\mu^*_\zeta(j)$ must remain above $\frac{1}{2k-j}$ for $\zeta \in (\xi, 1)$. □

10.5.3 The Average Positive Degree

The average positive degree of an endpoint partition (A, V) can be computed by dividing the total number of endpoints $|A| = m$ by the total number of vertices of degree 1 or greater, which is $\sum_{i>0} \Lambda_{(A,V)}(i)$. During the k-orientability process, the average positive degree is thus

$$\frac{m-t}{\sum_{i>0} \mathbf{\Lambda}_t(i)}.$$

Since we are scaling $\mathbf{\Lambda}_t(i)$ by dividing by the size of the original endpoint set, the average positive degree is given asymptotically by

$$\frac{1-\xi}{\sum_{i>0} \Lambda_\xi(i)}.$$

The following lemma confirms that, if the average positive degree is less than $2k$ and $\mathrm{dmin}_\xi > k$, then this quantity is decreasing, and therefore the asymptotic minimum degree must drop back down to k at some point before termination.

Lemma 10.5.4. *Assume that there exists $\xi < 1$ such that $\frac{1-\xi}{2k} < \sum_{i>0} \Lambda_\xi(i)$ and $\mathrm{dmin}_\xi \geq k+1$. Then there exists $\xi < \zeta < 1$ such that $\mathrm{dmin}_\zeta \leq k$.*

Proof. Since $\sum_{i=0}^{\infty} \Lambda_\xi(i)$ remains constant, we have $d\left(\sum_{i>0} \Lambda_\xi\right) = -d\Lambda_\xi(0)$. And, since all but $o(m)$ iterations of degree $j \geq k$ include $2k$ steps, and reduce $\mathbf{\Lambda}_t(0)$ by exactly 1, we have

$$d\Lambda_\xi(0) = d\sigma_\xi[L] = \frac{d\xi}{2k}$$

when $\mathrm{dmin}_\xi > k$.

It follows that

$$\frac{d}{d\xi}\left(\frac{1-\xi}{2k} - \sum_{i>0} \Lambda_\xi(i)\right) = 0$$

and hence the condition $\frac{1-\xi}{2k} < \sum_{i>0} \Lambda_\xi(i)$ continues to hold until dmin_ξ drops below k (or reaches $2k$).

Moreover, since the total number of endpoints (scaled by $1/m$) is given by

$$1 - \xi = \sum_{i>1} i\Lambda_\xi(i),$$

then clearly we must have $\mathrm{dmin}_\xi < 2k$ so long as $\frac{1-\xi}{2k} < \sum_{i>0} \Lambda_\xi(i)$. It follows that $\mathrm{dmin}_\xi < k$ must occur before $\mathrm{dmin}_\xi \geq 2k$.

Finally, note that, if $\mathrm{dmin}_\zeta \geq k$ holds for all $\zeta \in (\xi, 1)$, then for $\epsilon = \frac{1-\xi}{2k} - \sum_{i>0} \Lambda_\xi(i) > 0$, we would have

$$\sum_{i>0} \Lambda_{1-\epsilon}(i) = \frac{\epsilon}{2k} - \epsilon < 0,$$

which is a contradiction and therefore $\mathrm{dmin}_\zeta < k$ must occur for some $\zeta < 1 - \epsilon$. □

10.5.4 Proof of Theorem 10.5.1

The proof of theorem 10.5.1 is now relatively straightforward.

Proof of Theorem 10.5.1. By assumption, the $(k+1)$-core of degree $\boldsymbol{\mathcal{G}}(\mu)$ has average positive degree strictly less than $2k$. Since the k-orientability and the $(k+1)$-core processes coincide until $\mathbf{dmin}_t$ reaches $(k+1)$. This implies that the average positive degree at the first point ζ at which $\mathrm{dmin}_\zeta \geq k+1$ is strictly less than $2k$.

Hence, by lemma 10.5.4, there exists some $\xi > \zeta$ $\mathrm{dmin}_\xi < k$. Moreover, in order for such a ξ to exists, there must occur a transition from $\mathrm{dmin}_{\xi_0} = k+1$ to $\mathrm{dmin}_{\xi_0^+} = k$, at some time $\xi_0 < \xi$. Hence, by lemma 10.5.3, it follows that $\mathrm{dmin}_\xi \leq k$ for all $\xi \in (\zeta_0, 1)$. By proposition 10.3.15, this implies that $\boldsymbol{\mathcal{G}}(\mu)$ is w.p.h.p. k-orientable. □

10.6 k-Orientability Thresholds for $\mathbf{G}_{n,p}$

As discussed above, the k-orientability thresholds for the Erdős-Rényi random graphs $\mathbf{G}_{n,p}$ are particularly significant due to algorithmic applications. We now use the results of the previous section to compute these thresholds.

Recall that $\mathbf{G}_{n,p}$ with average degree $p = d/n$ can be simulated by the configuration model by specifying a Poisson degree distribution

$$\pi_d(i) = \frac{e^{-d}d^i}{i!},$$

and in this case the residual distribution will also be Poisson. Hence, it order to determine the k-orientability thresholds for $\mathbf{G}_{n,p}$, all that remains is to show that the Poisson distribution satisfies the conditions the Theorem 10.5.1.

Theorem 10.6.1. *For any $k > 1$, k-orientability threshold for the Erdős-Rényi random graph $\mathbf{G}_{n,p}$ coincides with the threshold at which the $(k+1)$ core has average positive degree $2k$.*

Proof. It suffices to show that the Poisson distribution satisfies the property that $\pi_d^*(i+1) \geq \pi_d^*(i)$ for all i, where

$$\pi_d^*(i) = \frac{\pi_d(i)}{\sum_{j \geq i} \pi_d(j)}.$$

Since $\pi_d(j) = \frac{e^{-d}d^j}{j!}$, we compute $\frac{\pi_d(j)}{\pi_d(i)} = \frac{i!}{j!}d^{j-i}$, and therefore

$$\frac{1}{\pi_d^*(i)} = \sum_{j=i}^{\infty} \frac{i!}{j!}d^{j-i} = i! \sum_{h=0}^{\infty} \frac{d^h}{(h+i)!}.$$

Hence

$$\frac{1}{\pi_d^*(i+1)} = (i+1)! \sum_{h=0}^{\infty} \frac{d^h}{(h+i+1)!} = i! \sum_{h=0}^{\infty} \left(\frac{i+1}{h+i+1}\right)\left(\frac{d^h}{(h+i)!}\right) \leq \frac{1}{\pi_d^*(i)}.$$

□

Using this theorem, it is not difficult to explicitly compute the k-orientability thresholds for $\mathbf{G}_{n,p}$; the values c_k for $k = \{2, 3, 4, 5\}$, alongside the thresholds for emergence of a $(k+1)$-core, are shown in table 10.1.

k	d_k	c_k
2	3.351	3.588
3	5.149	5.755
4	6.799	7.843
5	8.365	9.896

Table 10.1: The k-orientability threshold for $G_{n,d/n}$ for small values of k. In this table $d = c_k$ gives the k-orientability threshold we establish, and $d = d_k$ gives the threshold for the emergence of the $(k+1)$ core, which is the naive lower bound for k-orientability.

Chapter 11

Independent Sets in Random 3-regular Graphs

An independent set in a graph G is a subset of vertices, no two of which are adjacent, and the independence number of G is the size of a maximum independent set. Computing the independence number of a graph is a classical NP-hard optimization problem, in this section, we consider the the independent set problem for the special case of random 3-regular graphs.

The algorithm analysis of 3-regular graphs is easier in certain ways than for general random graphs, since the degree and residual distributions are trivial: we have $\lambda(3) = 1$ and $\mu(2) = 1$. This makes it possible to analyze more sophisticated algorithms, since we will not encounter infinite systems of differential equations; in fact, in many cases, the analysis can be reduced to a single differential equation in one variable. As such, 3-regular graphs are a useful testing ground for heuristics, as in many cases the limiting factor is our imagination rather than the algebraic difficulty of finding a solution.

For the independent set problem, the ease of analysis is counterbalanced by inherent combinatorial difficulty. In fact, unlike in the previous chapters, we will not attempt to determine the independence number with asymptotic precision; instead we will merely compute a lower bound on the independence number. For sparse random graphs, the independence number is generally of order $O(n)$, and hence the relevant quantity is the *independence ratio*, which is the ratio of the independence number to the number of vertices.

At present, the best lower and upper bounds for the independence ratio of random 3-regular graphs are $0.4328\ldots$ and $0.4615\ldots$, respectively. The lower bound was achieved by algorithmic analysis of a greedy heuristic by Frieze and Suen [36]. In this chapter, we will review this greedy heuristic, and reproduce this lower bound. Then, we present a slightly stronger heuristic, which achieves a lower bound of $0.43946\ldots$. In absolute terms, this

represents an improvement of just 0.0064... over the current bound; however, given that the distance between the lower and upper bounds is currently 0.0287..., this new bound reduces the gap by approximately 22%.

Chapter Organization

Since the analysis of our heuristic is similar to the analysis of the greedy heuristic, we first develop a general technique for deriving lower bounds in the independence ratio of random 3-regular graphs using heuristics. Specifically, theorem 11.1.1 in section 11.4 will allow us to translate certain attributes of an independent set heuristic to lower bounds more or less automatically.

We then proceed to analyze two particular heuristics using this framework. First, in section 11.2, we analyze the greedy heuristic, and in so doing we reproduce the result of Frieze and Suen. Then, in section 11.3, we introduce a more powerful heuristic which achieves the stronger lower bound. Finally, in section 11.4, we discuss possible generalizations of this technique, and we propose two open problems related to the limitations of greedy heuristics for finding independent sets in random regular graphs.

History and Background

As noted, the current lower bound for the independence ratio of a random 3-regular graph is due to an algorithmic analysis of the greedy heuristic by Frieze and Suen [36]. The upper bound of 0.4615 is due to Bollobás [12], and appears in his book [13]. Frieze and Suen cite a stronger upper bound of 0.4591, also attributed to the book of Bollobás , but this stronger bound does not appear in the copy of [13] owned by this author. To complicate matters, the newer edition of [13] mentions the result of Frieze and Suen, but as an upper bound (mistakenly) rather than a lower bound.

In any case, the potential controversy regarding the upper bound does not affect the present discussion, which deals with the lower bound. The methodology of this chapter is similar to that of Frieze and Suen, in that both involve algorithmic analysis of a heuristic. Our improved lower bound is due to a stronger heuristic; this result has yet been written

as an individual paper.

11.1 General Methodology

For any graph G, let us denote by $\Gamma(G)$ the size of the largest independent set in G; the ratio of $\Gamma(G)$ to the vertex set $|V|$ is the *independence ratio*, which we denote by $\gamma(G) = \Gamma(G)/|V|$. Our ultimate objective is to compute the independence ratio of a random 3-regular graph. However, this is a very difficult problem, and we will not actually determine this value precisely, and instead we will only attempt to compute lower bounds.

It is not known whether the independence ratio of a random 3-regular graph in fact converges w.e.h.p. (or even a.a.s.) to a constant, although it is widely believed that this is the case. Nevertheless, an asymptotic lower bound on the independence ratio can be defined as follows: letting $\boldsymbol{\mathcal{G}}(3)$ denote an asymptotic random 3-regular graph, and we define the quantity

$$\gamma = \sup\{q : \gamma(\boldsymbol{\mathcal{G}}(3)) \geq q \text{ w.e.h.p.}\}.$$

As noted above, it is not known, though widely conjectured, that γ is also an upper bound for the independence ratio $\gamma(\boldsymbol{\mathcal{G}}(3))$. Nevertheless, with some abuse of terminology, we will simply refer to γ without qualification as the w.e.h.p. independence ratio of a 3-regular graph. At present, the best lower and upper bounds for γ are $0.4328\ldots < \gamma < 0.4615\ldots$, and our objective in this chapter is to improve on this lower bound.

Our approach is to trace the execution of a simple heuristic on a random input graph. We thus define an *independent set (i-set) heuristic* to be a pair of (possibly random) mappings α, β from the set $\mathbb{G}$ of graphs to the sets $\mathbb{G}$ and $\mathbb{N}$ such that, for any G, we have

$$\Gamma(G) \geq \Gamma(\alpha(G)) + \beta(G). \tag{11.1}$$

The execution path of such a heuristic is thus a sequence (G_t) of graphs such that $G_{t+1} = \alpha(G_t)$, which terminates at the time τ when G_τ is an empty graph. And, such an execution

path will compute a lower bound on the independence number:

$$\Gamma(G) \geq \sum_{t=0}^{\tau-1} \beta(G_t).$$

A simple example of an i-set heuristic is the following: choose an arbitrary vertex v, add v to the independent set, and remove v and all of its neighbors from G to yield $\beta(G)$. In this case, we have $\beta(G) = 1$ for any (non-empty) graph, since exactly one vertex is added to the independent set each iteration. In general, though, it is not necessary to actually add vertices to construct an independent set explicitly in this way, as long as the inequality in (11.1) is satisfied.

As usual, we will work with heuristics which, when executed on a random input graph $\mathbf{G}$, preserve the uniform distribution of $\alpha(\mathbf{G})$, conditional on its endpoint partition (or, equivalently, on its degree sequence). In this case we may simply describe a heuristic in terms of its action on an endpoint partition. The mappings α and β are replaced by random mappings $\boldsymbol{\alpha} : \mathbb{H} \to \mathbb{H}$ and $\boldsymbol{\beta} : \mathbb{H} \to \mathbb{N}$, and the inequality in (11.1) relating the independence numbers of G to $\alpha(G)$ yields the following relationship between the distributions of the independence number of random graphs:

$$\Gamma(\boldsymbol{\mathcal{G}}(H)) \stackrel{d}{\triangleright} \Gamma(\boldsymbol{\mathcal{G}}(\boldsymbol{\alpha}(H))) + \boldsymbol{\beta}(H).$$

In this expression, the symbol $\stackrel{d}{\triangleright}$ denotes domination in distribution, meaning that for any i, the probability that $\Gamma(\boldsymbol{\mathcal{G}}(H)) \geq i$ is greater than or equal to the probability that $\Gamma(\boldsymbol{\mathcal{G}}(\boldsymbol{\alpha}(H))) + \boldsymbol{\beta}(H) \geq i$. Accordingly, by tracing the execution of such a heuristic, we may compute a lower bound in probability on $\Gamma(\boldsymbol{\mathcal{G}}(H))$.

11.1.1 Asymptotic Representation

As defined in chapter 3, the observable process corresponding to the execution of an i-set heuristic the a random sequence $(\mathbf{H}_t)$ of endpoint partitions $\mathbf{H}_t = (\mathbf{A}_t, \mathbf{V}_t)$, where $\mathbf{H}_{t+1} \stackrel{d}{=} \boldsymbol{\alpha}(\mathbf{H}_t)$. Since we are only considering 3-regular graphs, we will not make use of our general topological representation of an endpoint partition by its degree and residual distribution. Instead, we will make use of only four parameters:

1. the number of endpoints, which we denote by $\mathbf{m}_t$;

2. the number of non-empty vertices, which we denote by $\mathbf{n}_t$;

3. the number of vertices of degree 1, which we denote by $\mathbf{\Lambda}_t$;

4. the number of vertices currently added to the independent set, which we denote by $\mathbf{q}_t$.

Also, rather than counting the number of individual endpoint removals as we have done previously, we will instead simply keep track of the total number of iterations of a heuristic. Hence, the state $\mathbf{H}_t$ at time t is represents the endpoint partition resulting from t applications of the random mapping $\boldsymbol{\alpha} : \mathbb{H} \to \mathbb{H}$, and the first three parameters above are functions of $\mathbf{H}_t$.

The fourth parameter $\mathbf{q}_t$ is equal to the running total of the values $\boldsymbol{\beta}(\mathbf{H}_s)$ for $s < t$:

$$\mathbf{q}_t = \sum_{s=0}^{t-1} \boldsymbol{\beta}(\mathbf{H}_s)$$

As noted above, an i-set heuristic does not actually need to "add" vertices to the independent set; all that is necessary is that the inequality in (11.1) is satisfied. Accordingly, $\mathbf{q}_t$ can be more precisely defined as the amount by which the independence number of the original graph is guaranteed to exceed the independence number of graph after t iterations of the heuristic. Nevertheless, for simplicity of exposition, we will continue to refer to $\mathbf{q}_t$ as the current "size" of the independent set, rather than using more precise but less succinct terminology.

It is clear that the first three of these parameters completely specify an endpoint partition with maximum degree 3, modulo the number of empty vertices. During the execution of a heuristic, we will scale all four of these parameters, as well as the time coordinate, by dividing by the initial number of vertices $n = \mathbf{n}_0$, so a solution will satisfy $\mathbf{m}_{\lfloor \xi \mathbf{n} \rfloor}/\mathbf{n}_0 \to m_\xi$ and similarly for $\mathbf{n}_t$, $\mathbf{q}_t$, and $\mathbf{\Lambda}_t$. Note, then, that in the initial state, we have $n_0 = \mathbf{n}_0/\mathbf{n}_0 = 1$, $m_0 = \mathbf{m}_0/\mathbf{n}_0 = 3$, $\Lambda_0 =$ and $q_0 = 0$.

Unlike the algorithms we have analyzed previously, the duration of a heuristic is not specified exactly, since the number of vertices removed each iteration is itself a random variable, and therefore the termination time will also be random. We will represent this termination time again by scaling by the initial number of vertices, and hence the solution to $\boldsymbol{\tau}$ will be a constant τ such that $\boldsymbol{\tau}/n \to \tau$.

The number of vertices added to the independent set by such a heuristic will be the terminal value $\mathbf{q}_{\boldsymbol{\tau}}$. And, since $\mathbf{q}_{\boldsymbol{\tau}}$ is implicitly scaled by dividing by the initial number of endpoints, the independence ratio computed by such a heuristic is the terminal value q_τ of the solution to this process. It follows that the w.e.h.p. solution to q_τ for any such heuristic yields a lower bound on the independence ration γ of random 3-regular graphs.

Reduction to One Dimension

As described above, there are four parameters which we must take into account when analyzing an i-set heuristic. However, the asymptotic behavior of such a heuristic can in fact be described in terms of just one parameter as follows.

First, the number of vertices of degree 1 will remain $\mathbf{\Lambda}_t = o(n)$ w.e.h.p. throughout the execution of a the heuristic. This is because, if we ever encounter a vertex of degree 1, we may simply add this vertex to the independent set and remove its neighbor, and we may continue to do this until no vertices of degree 1 remain. Hence, vertices of degree 1 will never accumulate, and every w.e.h.p. solution will satisfy $\Lambda_\xi = 0$ for all ξ. Second, the current size of the independent set $\mathbf{q}_t$ does not affect the execution path of the algorithm, since the heuristic will behave the same way regardless current value of $\mathbf{q}_t$.

This leaves us with two relevant parameters: $\mathbf{m}_t$ and $\mathbf{n}_t$. However, the exact values of $\mathbf{m}_t$ and $\mathbf{n}_t$ do not generally affect the structure of the random graph; what matters is the the ratio of the two. Specifically, the average (positive) degree at time $t = \lfloor \xi n \rfloor$ is asymptotically given by m_ξ/n_ξ, and since the number of vertices of degree 1 is $o(n)$, then we have $2\lambda_\xi(2) + 3\lambda_\xi(3) = m_\xi/n_\xi$ and $\lambda_\xi(2) + \lambda_\xi(3) = 1$. It is therefore trivial to recover the degree distribution, as well as the residual distribution, from the ratio m_ξ/n_ξ of endpoints

to vertices.

For algebraic purposes, we will in fact work with the inverse of the average degree, which we denote by

$$\mathbf{z}_t = \mathbf{n}_t/\mathbf{m}_t.$$

This is because the residual distribution is more easily computed in terms of the inverse of the average degree: assuming that the number of vertices of degree 1 is exactly 0, then the number of endpoints of residual degree 1 (i.e. true degree 2) must be equal to

$$2(3\mathbf{n}_t - \mathbf{m}_t).$$

Hence, the residual distribution can be recovered by the parameter $z_\xi = n_\xi/m_\xi$ by

$$\mu_\xi(1) = 6z_\xi - 2$$
$$\mu_\xi(2) = 3 - 6z_\xi$$
$$M(\mu) = 4 - 6z_\xi.$$

11.1.2 General Solution to I-Set Heuristics

For the heuristics we will study, the expected increments of the three asymptotically "visible" parameters m_ξ, n_ξ, and q_ξ will be functions of the inverse average degree $z_\xi = n_\xi/m_\xi$. In order to invoke the differential equations method, we must therefore express the changes in these three parameters as functions of z_ξ as follows:

$$\begin{aligned} dm_\xi &= -f(z_\xi)d\xi \\ dn_\xi &= -g(z_\xi)d\xi \\ dq_\xi &= h(z_\xi)d\xi. \end{aligned} \tag{11.2}$$

Note that the sign is reversed in the first two of these equations; this is because both m_ξ and n_ξ decrease over time as vertices and endpoints are removed, while q_ξ increases as vertices are added to the independent set. Hence, all three of the functions f, g, h will be positive (or at least non-negative).

Since there are relatively few variables involved, we are able to find a general solution to the terminal value of q_τ of the size of the independent set found a particular heuristic in terms of the functions f, g, and h. This will free us to explore possible heuristics, and then apply a single formula to derive the corresponding bound $\gamma \geq q_\tau$ on the independence ratio of $\boldsymbol{\mathcal{G}}(3)$.

Theorem 11.1.1. *Assume there is a uniformity preserving i-set heuristic which satisfies the differential constrains in* (11.2)*, along with* $d\Lambda_\xi = 0$*, and assume further that:*

1. *for any* $\epsilon > 0$*, the functions* f*,* g*, and* h *are Lipschitz continuous for* $z \in [2+\epsilon, 3]$*;*
2. $zf(z) > g(z)$ *for all* $z \in [1/3, 1/2)$*.*

Then the w.e.h.p. independence ratio of a random 3*-regular graph is at least*

$$\gamma \geq \int_{z=1/3}^{1/2} \frac{3e^{-r(z)}h(z)/f(z)}{z - g(z)/f(z)} dz, \tag{11.3}$$

where

$$r(z) = \int_{y=1/3}^{z} \frac{1}{y - g(y)/f(y)} dy.$$

Proof. In order to solve the system (11.2), we first derive a differential constraint on $z_\xi = n_\xi/m_\xi$:

$$\begin{aligned} dz_\xi = d\left(\frac{n_\xi}{m_\xi}\right) &= \frac{-g(z_\xi)}{m_\xi} - \frac{-f(z)n_\xi}{m_\xi^2} \\ &= \frac{zf_\xi(z) - g(z_\xi)}{m_\xi} d\xi. \end{aligned}$$

Note that the condition $zf(z) > g(z)$ ensures that $dz_\xi > 0$ whenever $z_\xi < 1/2$. It follows that z_ξ will be strictly increasing, and we may change variables and express our solution as a function of $z \in [1, 3/2]$.

Since $n_\xi = z_\xi \cdot m_\xi$, then the relevant quantities are m_ξ and q_ξ, and we may compute the derivatives of these functions with respect to z_ξ by

$$\begin{aligned} \frac{dm_\xi}{dz_\xi} &= \frac{-f(z_\xi)m_\xi}{zf_\xi(z) - g(z_\xi)}, \\ \frac{dq_\xi}{dz_\xi} &= \frac{h(z_\xi)m_\xi}{zf_\xi(z) - g(z_\xi)}. \end{aligned}$$

To simplify our notation, we express m_ξ and q_ξ as functions $m(z)$ and $q(z)$ such that $m(z_\xi) = m_\xi$ and $q(z_\xi) = q_\xi$ for any time ξ. The above equations thus yield

$$m'(z) = \frac{-m(z)}{z - g(z)/f(z)} \tag{11.4}$$

and

$$q'(z) = \frac{m(z)h(z)/f(z)}{z - g(z)/f(z)}. \tag{11.5}$$

Note that the signs of the above derivatives are opposites: $q(z)$ increases with z, while $m(z)$ decreases with z. This is because, as noted above, we initially have $z_0 = 1/3$, $m_0 = m(1/3) = 3$, and $q_0 = q(1/3) = 0$, and as time passes both z_ξ and q_ξ increase, while m_ξ decreases. Of course, this is of little consequence, since we may reverse the signs and solve these equations in whatever direction we choose.

The Lipschitz conditions on f, g ensure that (11.4) has a unique solution satisfying the initial condition $m(1/3) = 3$, which extends to any interval $z \in [1/3, 1/2 - \epsilon]$, and this unique solution extends to all of extends to all of $z \in [1/3, 1/2)$. The graph of this solution represents the set of all possible pairs $(z_\xi, m_\xi) = (z_\xi, m(z_\xi))$ which occur in a solution to the process, since m_ξ is decreasing as a function of ξ, then the terminal state occurs when $m(z) = 0$.

The ordinary linear differential equation (11.4) with initial condition $m(1/3) = 3$ has solution $m(z) = 3e^{-r(z)}$, where

$$r(z) = \int_{y=1/3}^{z} \frac{1}{y - g(y)/f(y)} dy,$$

since in this case $m'(z) = -m(z)r'(z) = \frac{-m(z)}{z-g(z)/f(z)}$. Note also that $0 < r(z) < \infty$ for $z \in [1/3, 1/2)$, and it follows that $m(z) > 0$ for $z < 1/2$, so the terminal state must satisfy $z_\tau = 1/2$.

Now, it may be the case that $r(1/2) < \infty$, in which case $m(1/2) > 0$. In this situation, there will be a time τ_0 at which $z_{\tau_0} = 1/2$ and $m_{\tau_0} > 0$; and, this this time τ_0 will precede the actual termination time τ of the heuristic.

If this occurs, we cannot analyze the state for times $\xi > \tau_0$ using the variable z_ξ, since z_ξ remains constant during the interval $\xi \in [\tau_0, \tau]$. However, the integral

$$q(1/2) = \int_{z=1/3}^{1/2} q'(z) = \int_{z=1/3}^{1/2} \frac{m(z)h(z)/f(z)}{z - g(z)/f(z)} = \int_{z=1/3}^{1/2} \frac{3e^{-r(z)}h(z)/f(z)}{z - g(z)/f(z)} dz,$$

in equation (11.3) nevertheless yields the value q_{τ_0}, which represents the number of vertices added to the independent set at the time τ_0 when z_{τ_0} reaches $1/2$. And, since this $q_\tau \geq q_{\tau_0}$, this yields a lower bound on γ. For the heuristics we analyze, we will have $\tau_0 = \tau$, and therefore the above integral will in fact compute the value of q_τ correctly, which yields the best possible bound on γ achievable by that particular algorithm. □

For any given heuristic, this theorem yields an essentially automated method for computing the corresponding lower bound on γ, assuming that we are able to compute the functions f, g, and h in (11.2). Moreover, since we have already solved the relevant differential equation in the proof of this theorem, then the bound on γ in equation (11.3) can be found just by computing two definite integrals. While this may not be "easy" in all cases, it is certainly a tractable problem, since even if the functions involved are particularly unpleasant, we may compute a solution numerically.

On the other hand, this theorem does not yield a simple intuitive interpretation of the resulting bound on γ. Hence, we now give a slightly different account of this formula. For any $z \in [1/3, 1/2]$, let us denote by H_z an asymptotic endpoint partition with minimum degree 2, maximum degree 2, and inverse average degree $n/m \to z$, so H_z corresponds to the state of an i-set heuristic when $z_\xi = n_\xi/m_\xi = z$.

For such an endpoint partition, let us define

$$\gamma(z) = \sup\{q : \gamma(\boldsymbol{\mathcal{G}}(H_z)) \geq q\}.$$

That this function is well-defined, and $\gamma = \gamma(1/3)$ is the independence ratio of a random 3-regular graph. At this point we will not attempt to rigorously establish any non-trivial properties of this function. However, the corollary to theorem (11.1.1) we prove below implies that $\gamma(z)$ is strictly decreasing; intuitively, this is because a lower value of z corresponds to a higher average degree, which would suggest a lower independence ratio.

It does not appear difficult to argue that $\gamma(z)$ must be also be continuous, but this is not essential to our analysis, and hence we will not attempt a proof. On the other hand, it not intuitively clear whether or not $\gamma(z)$ is differentiable (although, if $\gamma(z)$ is monotonic, then it must be differentiable outside a null set). A "best guess" is that $\gamma(z)$ is at least piecewise differentiable, though perhaps not continuously differentiable on all of $[1/3, 1/2]$.

In any case, we may speak of the derivative of $\gamma(z)$ for the purpose of establishing lower bounds; for instance, for given $y \in (1/3, 1/2)$, a bound of $\gamma'(y) \geq c$ indicates that

$$\liminf_{\epsilon \to 0} \frac{\gamma'(y+\epsilon) - \gamma'(y)}{\epsilon} \geq c.$$

Theorem 11.1.1 can be understood in terms of the derivative of $\gamma(z)$ in the following manner.

Corollary 11.1.2. *Given the conditions of Theorem 11.1.1, the function $\gamma(z)$ satisfies*

$$z^2 \gamma'(z) \leq \frac{\gamma(z) - h(z)/g(z)}{f(z)/g(z) - 1/z}. \tag{11.6}$$

for all $z \in (1/3, 1/2)$

Proof. Consider executing this heuristic such that $\mathbf{z}_0 = y$ for some fixed $y \in (1/3, 1/2)$, and stopping at after ϵn steps for some small but fixed ϵ. At this time, the (scaled) number of vertices, the inverse average degree, and the (scaled) size of the independent set will be given asymptotically by n_ϵ, z_ϵ, and q_ϵ, respectively. Also, the execution of this heuristic will satisfy the differential constraints in (11.2).

In particular, the independence ratio of the graph which remains at termination is asymptotically (at least) $\gamma(z_\epsilon)$ w.e.h.p. by assumption, and since there are n_ϵ vertices remaining, then there will be a total of $\gamma(z_\epsilon) n_\epsilon$ vertices in such a maximum independent set. Hence, by combining this independent set with the number of vertices already added by our heuristic, the independence ratio of our original graph is at least

$$q_\epsilon + \gamma(z_\epsilon) n_\epsilon,$$

This is clearly a lower bound on $\gamma(y)$, and by taking the limit as $\epsilon \to 0$, we have

$$\lim_{\epsilon \to 0} \frac{\gamma(y) - \gamma(z_\epsilon) n_\epsilon - q_\epsilon}{\epsilon} \geq 0. \tag{11.7}$$

Since $z_0 = y$, $n_0 = 1$, and $m_0 = n_0/y$ by assumption, then as in the proof of the Theorem 11.1.1, we have

$$\lim_{\epsilon \to 0} \frac{z_\epsilon - y}{\epsilon} = \frac{yf(y) - g(y)}{m_0} = y^2(f(y) - g(y)/y)$$

and this yields an inequality

$$-\gamma'(y)y^2(f(y) - g(y)/y) + g(y)\gamma(y) - h(y) \geq 0$$

from which we may compute

$$\begin{aligned} y^2\gamma'(y) &\leq \frac{g(y)\gamma(y) - h(y)}{yf(y) - g(y)} \\ &= \frac{\gamma(y) - h(y)/g(y)}{f(y)/g(y) - 1/y}. \end{aligned}$$

□

In order to make use of the differential equation (11.6) to compute $\gamma = \gamma(1/3)$, we must know the value of $\gamma(z)$ for some $z \in [1/3, 1/2]$. Although we omit the proof, it is not difficult to see that $\gamma(1/2) = 1/2$. This is because if $z = m/n = 1/2$, then every vertex in the graph must have degree 2. The structure of such a graph is the disjoint union of cycles, and it can be shown that that all but $o(n)$ vertices will belong to a cycle of length $\Omega(n)$. And, since a cycle of length j has independence number $\lfloor j/2 \rfloor$, then approximately one half of the vertices in such a long cycle can be placed into an independent set.

The value of γ can therefore be computed by solving (11.6) with initial condition $\gamma(1/2) = 1/2$; in fact, the general solution for $\gamma = \gamma(1/3)$ to this equation will coincide with the value computed in Theorem 11.1.1.[1] However, since we already have an explicit solution, we will not make use of this alternate (and equivalent) method of computing γ except for intuitive purposes.

[1]There is one case where the solution to (11.6) and the value computed in (11.3) do not coincide. This occurs if the function $r(z)$ in (11.3) does not satisfy $r(1/2) = \infty$. In this situation, the heuristic will reach average degree $1/z_{\tau_0} = 2$ before termination, and (11.3) does not take into account vertices added to the independent set during the interval $[\tau_0, \tau]$ when $z_\tau = 1/2$. Hence, in this case, (11.6) will yield a stronger bound than (11.3).

Intuitively, the differential equation (11.6) can be understood as follows. First, a smaller value of $\gamma'(z)$ corresponds to a larger independent set, since $\gamma(1/2) = 1/2$, and $\gamma(1/3) = 1/2 - \int_{z=1/3}^{1/2} \gamma(z)$. Hence, a better i-set heuristic will yield a lower value of $\gamma'(z)$ as computed above.

There are two relevant quantities in equation (11.6): the ratios $h(z)/g(z)$, and $f(z)/g(z)$. The value $h(z)/g(z) = dq_\xi/dn_\xi$ represents the rate at which vertices are added to the independent set relative to the rate at which vertices are being removed from the graph. So, if the value of $h(z)/g(z)$ were to remain constant for all z, then the size of the resulting independent set would be precisely $h(z)/g(z)$.

However, $h(z)/g(z)$ does not remain constant: as z increases, the average degree $m/n = 1/z$ decreases, and hence it becomes easier to fit more vertices in an independent set. As a consequence, for any reasonable heuristic, the ratio $h(z)/g(z)$ will increase with z.

The ratio $f(z)/g(z)$ in the denominator of (11.6) represents the average number of endpoints which are removed from the graph for every vertex removed. And, the larger the value of $f(z)/g(z)$, the more rapidly the average degree $m/n = 1/z$ will decrease. It is therefore preferable to make the ratio $f(z)/g(z)$ as large as possible, since in this case, the algorithm will spend less time in the more difficult regions of the parameter space where the average degree m/n is large.

Informally, then, an ideal heuristic will maximize both $h(z)/g(z)$ and $f(z)/g(z)$, so as to increase the size of the independent set quickly, while simultaneously reducing the average degree as quickly as possible. Note, though, that the value of $\gamma(z)$ also appears on the right-hand side of (11.6), and therefore the ideal trade-off between $h(z)/g(z)$ and $f(z)/g(z)$ will depend on the actual size of the maximal independent set. Of course, we do not know the value of $\gamma(z)$; but, if we are given a lower bound on $\gamma(z)$, this equation can tell us immediately, based on the values of $f(z)/g(z)$ and $h(z)/g(z)$, whether a given heuristic will improve upon this bound.

11.1.3 Computing Differential Constraints

We now describe how to compute the functions f, g, and h in equation (11.2) for a particular i-set heuristic. First, recall that, although there are only 3 asymptotically visible parameters, there is one additional parameter $\mathbf{\Lambda}_t$, which counts the number of vertices of degree 1. And, although $\mathbf{\Lambda}_t$ remains $o(n)$ throughout the execution of a heuristic, we must nevertheless take vertices of degree 1 into account when computing the differential equations which govern the execution.

In order to apply the differential equations method, these four parameters must satisfy Lipschitz conditions. Also, the heuristic must preserve the conditional uniformity when executed on a random input graph. Accordingly, we shall say an i-set heuristic is *regular* if:

1. conditional uniformity is preserved;
2. the increments of $\mathbf{m}_t$, $\mathbf{n}_t$, $\mathbf{q}_t$, and $\mathbf{\Lambda}_t$ are bounded by a constant C.

For a regular i-set heuristic, expected increments of these four parameters can be translated into w.e.h.p. differential constraints using martingale concentration as usual. In particular, if a regular i-set heuristic does not affect the number of vertices of degree 1 (i.e. so $\Delta\mathbf{\Lambda}_t = 0$), then we may use the expected increments of of the other three parameters to compute these differential constraints as follows.

Proposition 11.1.3. *Assume the expected increments of a regular i-set heuristic are given by*

$$\begin{aligned}
\mathsf{E}_t[\Delta\mathbf{m}_{t+1}] &= -f(\mathbf{z}_t) \\
\mathsf{E}_t[\Delta\mathbf{n}_{t+1}] &= -g(\mathbf{z}_t) \\
\mathsf{E}_t[\Delta\mathbf{q}_{t+1}] &= h(\mathbf{z}_t) \\
\mathsf{E}_t[\Delta\mathbf{\Lambda}_{t+1}] &= 0
\end{aligned}$$

for continuous functions f, g, and h, and within an additive factor of $\pm o(1)$.

Then the differential constraints in (11.2) *hold w.e.h.p.*

Proof. In this case we have the differential constraint $d\Lambda_\xi = 0$ and therefore $\Lambda_\xi = 0$ w.e.h.p. for all ξ. Hence, the only parameters which change non-negligibly throughout the process are $\mathbf{m}_t$, $\mathbf{n}_t$, and $\mathbf{q}_t$, which will satisfy the differential constraints in (11.2) by martingale concentration. □

Proposition 11.1.3 is of course only applicable to heuristics which do not affect the number of vertices of degree 1. However, it is difficult to directly create heuristics which satisfy this property; indeed, anytime we remove a random endpoint, the probability of creating a vertex of degree 1 is equal to probability $\boldsymbol{\mu}_t(1)$ of choosing an endpoint of residual degree 1.

As mentioned above, these vertices of degree 1 can be greedily added to the independent set, so $\mathbf{\Lambda}_t$ never grows larger than $o(n)$. The procedure vertices of degree 1 are present is the following:

1. choose any vertex v of degree 1;
2. remove v along with its neighbor u;
3. remove all edges incident on the neighboring vertex u.

We shall call this the *greedy degree*-1 heuristic. It is clear that conditional uniformity is preserved by this heuristic, since we are simply removing the edge incident on the chosen vertex v. We now compute the expected increments of our four parameters during one iteration of this greedy heuristic.

Proposition 11.1.4. *The greedy degree*-1 *heuristic produces expected changes of*

$$\begin{aligned} \mathsf{E}_t[\Delta\mathbf{m}_{t+1}] &= -f_1(\mathbf{z}_t) & f_1(z) &= 2(5-6z) \\ \mathsf{E}_t[\Delta\mathbf{n}_{t+1}] &= -g_1(\mathbf{z}_t) & g_1(z) &= -2 \\ \mathsf{E}_t[\Delta\mathbf{q}_{t+1}] &= h_1(\mathbf{z}_t) & h_1(z) &= 1 \\ \mathsf{E}_t[\Delta\mathbf{\Lambda}_{t+1}] &= -k_1(\mathbf{z}_t) & k_1(z) &= (3-6z)^2 \end{aligned}$$

within an additive factor of $\pm o(1)$*, and assuming that* $0 < \mathbf{\Lambda}_t = o(n)$.

Proof. It is trivial that exactly 1 vertex is added to the independent set. Also, with probability $1 - o(1)$, exactly 2 vertices are removed: the selected vertex v of degree 1, and its neighbor. An additional vertex may be removed if one of the randomly selected endpoints happens to belong to a vertex of degree 1, but this occurs with probability $o(1)$, and does not contribute to the asymptotic estimate.

We may compute the number of endpoints removed as follows. First, we remove the single endpoint belonging to v along with its match. Then, we remove all endpoints belonging to the neighbor of v, along with their matches. The expect number of additional endpoints belonging to the neighbor of v is asymptotically $M(\mu) = 4 - 6z$. And, the probability that any of these endpoints match to each other is $o(1)$, so each one accounts for two additional expected endpoint removals, for a total of

$$2 + 2M(\mu) = 10 - 12z.$$

For the number of vertices of degree 1, clearly one such vertex is removed initially. Additional vertices of degree 1 may be created during the random removals, since with probability $\mu(1) = 6z - 2$, a randomly selected endpoint will produce a new vertex of degree 1. There are $M(\mu)$ such random selections in expectation, so we compute

$$\begin{aligned} M(\mu)\mu(1) - 1 = (4 - 6z)(6z - 2) - 1 &= 36z^2 + 36z - 9 \\ &= -(3 - 6z)^2. \end{aligned}$$

□

It is not difficult to see that this heuristic is optimal, meaning that the independence number of any graph G which contains a vertex v of degree 1 is exactly 1 greater then the independence number of the graph G' with this vertex and its neighbor removed. We will therefore employ the greedy degree-1 heuristic described above above whenever vertices of degree 1 are present.

Now, since the maximum degree is 3, then there are only two other possibilities for the minimum degree. Moreover, minimum degree 3 occurs only at the very beginning of

the algorithm, since we can only have minimum degree 3 if $m_\xi/n_\xi = 1/z_\xi = 3$, and this only occurs for $\xi = 0$, assuming that $dz_\xi > 0$. Accordingly, an i-set heuristic of this kind can be completely specified by the behavior when the minimum degree is exactly 2.

Proposition 11.1.5. *Assume a regular i-set heuristic satisfies expected increments*

$$\begin{aligned}
\mathsf{E}_t[\Delta \mathbf{m}_{t+1}] &= -f^*(\mathbf{z}_t) \\
\mathsf{E}_t[\Delta \mathbf{n}_{t+1}] &= -g^*(\mathbf{z}_t) \\
\mathsf{E}_t[\Delta \mathbf{q}_{t+1}] &= h^*(\mathbf{z}_t) \\
\mathsf{E}_t[\Delta \mathbf{\Lambda}_{t+1}] &= k^*(\mathbf{z}_t)
\end{aligned}$$

whenever the minimum degree is exactly 2*, within an additive factor of* $\pm o(1)$*, and such that the functions* f^*, g^*, h^*, k^* *are continuous.*

*Then execution of this heuristic, combined with the greedy degree-*1 *heuristic whenever* $\mathbf{\Lambda}_t > 0$*, satisfies the w.e.h.p. differential constraints*

$$\begin{aligned}
dm_\xi &= -\frac{f(z_\xi)}{1 + j(z_\xi)} \qquad & f(z) &= f^*(z) + g_1(z)j(z) \\
dn_\xi &= -\frac{g(z_\xi)}{1 + j(z_\xi)} \qquad & g(z) &= g^*(z) + g_1(z)j(z) \\
dq_\xi &= \frac{h(z_\xi)}{1 + j(z_\xi)} \qquad & h(z) &= h^*(z) + g_1(z)j(z),
\end{aligned} \tag{11.8}$$

where

$$j(z) = k^*(z)/k_1(z) = k^*(z)/(3 - 6z)^2.$$

Proof. Let us denote by $\sigma_t[\mathbf{L}(1)]$ and $\sigma_t[\mathbf{L}(2)]$ the total number of iterations in the time interval $\{1, \ldots, t\}$ for which the minimum degree is 1 and 2, respectively. Since each iteration of degree 2 increases $\mathbf{\Lambda}_t$ by $k^*(\mathbf{z}_t)$ in expectation, while each iteration of degree 1 increases $\mathbf{\Lambda}_t$ by $k_1(\mathbf{z}_t)$, we have the differential constraint

$$d\Lambda_\xi = k^*(z_\xi)d\sigma_\xi[L(2)] - k_1(z_\xi)d\sigma_\xi[L(1)].$$

Also, we have $\Lambda_\xi = 0$ for all ξ, and therefore $d\Lambda_\xi = 0$, so we may compute

$$d\sigma_\xi[L(1)] = d\sigma_\xi[L(2)]k^*(z_\xi)/k_1(z_\xi) = d\sigma_\xi[L(2)]j(z_\xi).$$

It follows that each iteration of the given heuristic will be followed in expectation by $j(z_\xi)$ iterations of the greedy degree-1 heuristic. We may therefore compute $d\sigma_\xi[L(1)] = \frac{1}{1+j(z_\xi)}d_\xi$ and $d\sigma_\xi[L(2)] = \frac{j(z_\xi)}{1+j(z_\xi)}d\xi$, hence the differential constraints on the parameter m_ξ can be computed by

$$\begin{aligned} dm_\xi &= -f^*(z_\xi)d\sigma_\xi[L(2)] - f_1(z_\xi)d\sigma_\xi[L(1)] \\ &= -\frac{f^*(z_\xi) - f_1(z_\xi)j(z_\xi)}{1+j(z_\xi)} \\ &= -\frac{f(z_\xi)}{1+j(z_\xi)}. \end{aligned}$$

And, the differential constraints on n_ξ and q_ξ can be computed identically. □

Note that the differential constraints in (11.8) are not exactly identical to those in (11.2) which were used in Theorem 11.1.1. However, the lower bound on γ computed in (11.3) only depends on the fractions $g(z)/f(z)$ and $h(z)/f(z)$, and therefore the factor of $(1+j(z))$ in the denominator of the three expressions in (11.8) will cancel. Hence the bound in Theorem 11.1.1 can be computed directly using the functions f, g, and h defined above.

11.2 The Greedy Heuristic

We now use the machinery from the previous chapter to reproduce the result of Frieze and Suen, which establishes a lower bound on the independence ratio γ of a random 3-regular graph by tracing the execution of the greedy heuristic.

This heuristic is quite simple:

1. choose a vertex v of minimum degree:
2. if there are any loops incident on v, simply remove v from the graph, along with all incident edges;
3. otherwise, add v to the independent set, and remove v from the graph along with all of its neighbors.

By analyzing the executing of this heuristic on a random input graph, we will achieve the following bound.

Theorem 11.2.1. *The greedy heuristic establishes a bound of*

$$\gamma \geq 6\ln(3/2) - 2 = 0.43279\ldots$$

on the independence ratio of a random 3*-regular graph.*

As with the greedy degree-1 heuristic, it is clear that this heuristic preserves conditional uniformity, since every edge exposed by the heuristic is subsequently removed, and hence the remaining endpoints will be matched uniformly at random. In order to invoke Theorem 11.1.1, we must therefore compute the expected increments of the four parameters.

Lemma 11.2.2. *The greedy heuristic satisfies*

$$\begin{aligned}
\mathsf{E}_t[\Delta\mathbf{m}_{t+1}] &= -f^*(\mathbf{z}_t) & f^*(z) &= 2f_1(z) = 4(5-6z)\\
\mathsf{E}_t[\Delta\mathbf{n}_{t+1}] &= -g^*(\mathbf{z}_t) & g^*(z) &= 3\\
\mathsf{E}_t[\Delta\mathbf{q}_{t+1}] &= h^*(\mathbf{z}_t) & h^*(z) &= 1\\
\mathsf{E}_t[\Delta\mathbf{\Lambda}_{t+1}] &= k^*(\mathbf{z}_t) & k^*(z) &= 2-2k_1(z) = 2-(3-6z)^2
\end{aligned}$$

within an additive factor of $o(1)$*, whenever* $\Lambda_t = 0$*.*

Proof. The analysis is similar to the greedy degree-1 heuristic. First, unless the two endpoints belonging to v match to each other (which occurs with probability $o(1)$), one vertex is added to the independent set and exactly three vertices are removed: the selected vertex v and its two neighbors. We thus have $h^*(z) = 1$ and $g^*(z) = 3$.

For the number of endpoints removed, we first count 4 endpoints (again with probability $1-o(1)$) for the two edges incident on v. And, since each of the matched vertices has expected residual degree $M(\mu) = 4-6z$, then expected number of additional edges removed is approximately $2(4-6z)$, for a total of

$$f^*(z) = 4 + 4(4-6z) = 20 - 24z = 2f_1(z)$$

expected endpoint removals, within an additive factor of $\pm o(1)$.

Also, for each of these $2M(\mu)$ edges removed, a new vertex if degree 1 is created if the matched endpoint has residual degree 1. This occurs with probability $\mu(1) = (6z - 2)$ for each edge, and hence in expectation, each iteration creates

$$\begin{aligned} k^*(z) = 2M(\mu)\mu(1) = 2(4 - 6z)(6z - 2) &= 2(-36z^2 + 36z - 8) \\ &= 2 - (3 - 6z)^2 \\ &= 2 - 2k_1(z) \end{aligned}$$

new vertices of degree 1 (again within $\pm o(1)$). □

It is now routine, if somewhat tedious, to compute the lower bound on γ.

Proof of Theorem 11.2.1. In order to invoke Theorem 11.1.1, it is first necessary compute the fractions $g(z)/f(z)$ and $h(z)/f(z)$ for the functions functions f, g, and h from equation (11.8) of proposition 11.1.5

Since $j(z) = k^*(z)/k_1(z)$, and $g_1(z) = 2$, we may directly compute

$$\begin{aligned} g(z)/f(z) = \frac{g^*(z) + g_1(z)j(z)}{f^*(z) + f_1(z)j(z)} &= \frac{k_1(z)g^*(z) + k^*(z)g_1(z)}{k_1(z)f^*(z) + k^*(z)f_1(z)} \\ &= \frac{3k_1(z) + 2(2 - 2k_1(z))}{(2k_1(z) + 2 - 2k_1(z))f_1(z)} \\ &= \frac{4 - (3 - 6z)^2}{4(5 - 6z)} \\ &= \frac{-36z^2 + 36z - 5}{4(5 - 6z)} \\ &= \frac{(5 - 6z)(6z - 1)}{4(5 - 6z)} \\ &= \frac{6z - 1}{4}. \end{aligned}$$

Therefore $\frac{1}{z - g(z)/f(z)} = \frac{4}{1 - 2z}$, and we may easily integrate

$$r(z) = \int_{y=1/3}^{z} \frac{1}{y - g(y)/f(y)} dy = -2(\ln(1 - 2z) + \ln 3),$$

to compute

$$m(z) = 3e^{-r(z)} = 27(1-2z)^2 = 3(3-6z)^2.$$

For a "sanity check," we may confirm that $m(1/3) = 3$ and $m(1/2) = 0$.

Next, we similarly compute

$$\begin{aligned} h(z)/f(z) = \frac{k_1(z)h^*(z) + k^*(z)h_1(z)}{k_1(z)f^*(z) + k^*(z)f_1(z)} &= \frac{k_1(z) + (2 - 2k_1(z))}{2f_1(z)}, \\ &= \frac{2 - (3-6z)^2}{4(5-6z)}. \end{aligned}$$

We now have

$$\frac{m(z)}{z - g(z)/f(z)} = \frac{3(3-6z)^2}{(1-2z)/4} = 36(3-6z)$$

and therefore

$$\begin{aligned} \frac{m(z)h(z)/f(z)}{z - g(z)/f(z)} &= \left(\frac{36}{4}\right)\left(\frac{(3-6z)(2-(3-6z)^2)}{5-6z}\right) \\ &= 9\left(\frac{2(3-6z) - (3-6z)^3}{5-6z}\right) \\ &= 9\left(\frac{(6z)^3 - 9(6z)^2 + 25(6z) - 21}{5-6z}\right) \\ &= 9\left(-(5-6z)^2 + 6(5-6z) - 10 + \frac{4}{5-6z}\right). \end{aligned}$$

The bound on γ is the integral of this function from $z = 1/3$ to $z = 1/2$:

$$\begin{aligned} \gamma &\geq 9\int_{z=1/3}^{1/2} \frac{m(z)h(z)/f(z)}{z - g(z)/f(z)} \\ &= 9\left(\frac{8-27}{18} - \frac{6(4-9)}{12} - 10/6 - \frac{\ln 2 - \ln 3}{6}\right) \\ &= 6\ln(3/2) - 2. \end{aligned}$$

□

11.3 An Improved Heuristic

We now introduce a slightly modified greedy heuristic which achieves a better lower bound on the independence ratio γ of 3-regular graphs. The basic idea is as follows. The

greedy heuristic defined above simply chooses a vertex v of minimum degree and adds it to the independent set. Now, if the minimum degree is 1, then this operation is optimal and cannot be improved. However, if the minimum degree is 2, then adding v to the independent set may yield a sub-optimal independent set.

Ideally, then, when the minimum degree is 2, we would prefer to incorporate some additional information into the decision of whether or not v belongs in our independent set. One possible way to accomplish this is to explore a small neighborhood of v before making our decision. Unfortunately, if we do this, we risk violating the conditional uniformity of the unexposed portion of the graph, which complicates the analysis of the algorithm.

However, note that, once we add v to our independent set, a small neighborhood of v becomes exposed in any case, since both edges incident on v, as well as all of the edges incident on the neighbors of v are exposed and then removed. Accordingly, since this information is going to be revealed regardless, we might as well expose these edges *before* making a decision about v, rather than afterwords, when it is too late to make use of this information. The question, then, is how this information can be used to make a better decision about v.

Our modified greedy heuristic will only take into account two pieces of information in addition to the degree of v: the degrees of both of the neighbors of v. Hence, before deciding what to do about v, we first expose both edges incident on v. Now, if both of these neighbors have degree 3, we will simply add v to the independent set as in the greedy algorithm.

However, if at least one of these neighbors also has degree 2, then the decision to add v to the independent set is somewhat arbitrary and premature; indeed, if u and v are adjacent, and both have degree 2, then there is no a priori reason to add v to the independent set rather than u. Hence, rather than simply adding v to the independent set in this case, we will proceed in a slightly different manner, based on the following observation about maximum independent sets in such a situation.

Proposition 11.3.1. *Let G be a graph with two adjacent vertices u, v of degree 2, and let*

u_1, v_1 denote the other neighbors of v and v, respectively. Then

$$\Gamma(G) = \Gamma(G') + 1,$$

where G' is the induced sub-graph of G on the vertex set $V \backslash \{u, v\}$, along with an edge connecting u_1 and v_1.

Proof. We first prove that $\Gamma(G) \geq \Gamma(G') + 1$, so let W' be any maximum independent set in G'. Since u_1 and v_1 are connected by an edge, then at most one these vertices belong to W'. W.l.o.g., we may assume that $v_1 \notin W'$, and in this case the set $W = W \cup \{v\}$ is an independent set in G of size $|W'| + 1 = \Gamma(G') + 1$.

For the reverse implication, it suffices to prove that there exists a maximum independent W set of G which contains either u or v, since in this case, only one of u_1 or v_1 can belong to this independent set, and therefore the same vertex subset yields an independent set of size $|W| - 1$ in G'. To prove this let W be any independent set of G which does not contain either u or v. We may create a new independent set if size at least $|W|$ by adding the vertex v and removing the vertex v_1 (if in fact $v_1 \in W$). This new independent set will have the same size as W, and will contain v, and the proof is complete. □

Accordingly, when the minimum degree is 2, our modified greedy heuristic will choose a vertex v of minimum degree and proceed as follows:

1. if the two endpoints belonging to v are matched to each other, simply remove v from the graph;
2. if both neighbors have degree 3, then add v to the independent set, and remove v, both neighbors, and all incident edges;
3. if at least one neighbor u has degree 2:
   (a) remove u and v, and the edge connecting them;
   (b) create a new edge connecting the neighbors u_1 and v_1 of u and v;

(c) increase the independence number by 1.

Note that, unlike in the greedy heuristic, we are not actually adding any vertices to the independent set of one of the neighbors has degree 1. We are simply establishing a bound relating the independence numbers of G and the modified graph G'. Intuitively, this heuristic somewhat resembles the k-orientability heuristic from the previous chapter, since we are essentially postponing the decision as to which vertex belongs to the independent set until after the recursive call to the modified graph G' has completed.

Also, note that, as with the greedy degree-1 heuristic, the modified heuristic above is optimal unless the vertex v has two neighbors of degree 3. Informally, then, this heuristic is strictly superior to the greedy heuristic, since in the only situation in which these heuristics differ is when v has at least one neighbor of degree 2, in which case the modified heuristic behaves optimally, while the greedy heuristic does not. Accordingly, it is perhaps expected that this heuristic achieves a stronger lower bound on γ.

Theorem 11.3.2. *The modified greedy heuristic establishes a bound of*

$$\gamma \geq \frac{1181}{2688} = 0.43936\ldots$$

on the independence ratio of a random 3*-regular graph.*

Unlike for the greedy heuristic, it is not entirely trivial that conditional uniformity is preserved, since we are no longer simply removing edges as they become exposed. Hence, before computing expected increments, we first establish conditional uniformity.

Lemma 11.3.3. *The modified greedy heuristic preserves conditional uniformity.*

Proof. The proof is similar to the proof of conditional uniformity for the k-orientation heuristic from the previous chapter. Hence, let $\mathbf{v}$ denote the vertex of degree 2 chosen by the heuristic, and we consider the three possible outcomes of a single iteration of the heuristic. Now, clearly, if a loop is found on $\mathbf{v}$, then removing this loop preserves conditional uniformity; also, if both neighbors have degree 3, then we are simply removing exposed edges from the graph as usual, and hence in this case conditional uniformity is maintained as well.

We must therefore deal with the third possible outcome, which occurs when at least one neighbor has degree 2. We denote the relevant vertices and endpoints at this point as follows:

- $\mathbf{u}$ denotes the selected neighbor of $\mathbf{v}$ with degree 2;
- $\mathbf{u}_1$ and $\mathbf{v}_1$ denote the additional neighbors of $\mathbf{u}$, and $\mathbf{v}$, respectively;
- $\mathbf{a}_0$ and $\mathbf{a}_1$ denote the (unexposed) endpoints on $\mathbf{u}$ and $\mathbf{u}_1$ which match to each other;
- $\mathbf{b}_1$ denotes the (exposed) endpoint on $\mathbf{v}_1$ which matches to $\mathbf{v}$.

Note that both neighbors of $\mathbf{v}$ may have degree 2; in this case we may arbitrarily designate one of these as the vertex $\mathbf{u}$ which will be removed by the heuristic.

Also, in order to determine the degrees of the neighbors of $\mathbf{v}$, at this point we have already exposed both edges incident on $\mathbf{v}$. However, the endpoints $\mathbf{a}_0$ and $\mathbf{a}_1$ which connect $\mathbf{u}$ to its neighbor $\mathbf{u}_1$ remain unexposed at this time. And, while our state of information includes the identity of the vertices $\mathbf{u}$, $\mathbf{u}_1$, and $\mathbf{v}_1$, as well as the endpoint $\mathbf{a}_0$, the vertex $\mathbf{u}_1$ and the endpoint $\mathbf{a}_1$ remain unknown. The the state after the two edges incident on $\mathbf{v}$ have been exposed, but previous to making structural modifications, is illustrated in figure 11.1.

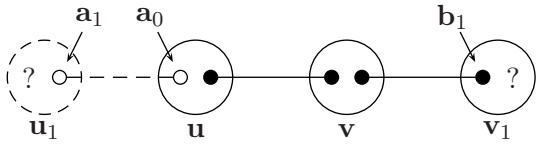


Figure 11.1: The state of the modified greedy heuristic previous to structural modifications; the edge connecting $\mathbf{a}_0$ and $\mathbf{a}_1$ remains unexposed at this time, and the identity of the vertex $\mathbf{u}_1$ is unknown.

At this point the heuristic removes the vertices $\mathbf{u}$ and $\mathbf{v}$, and creates a new edge connecting $\mathbf{v}_1$ and $\mathbf{v}_2$. Now, one way to accomplish this is to simply create a new edge connecting the endpoints $\mathbf{a}_1$ and $\mathbf{b}_1$. However, since we do not know the identity of $\mathbf{a}_1$ at this time, then this method

In order to preserve conditional uniformity, though, we may perform these modifications by replacing the exposed endpoint $\mathbf{b}_1$ on $\mathbf{v}_1$, with the unexposed endpoint $\mathbf{a}_0$ on $\mathbf{u}$, and in this way the edge $\{\mathbf{a}_0, \mathbf{a}_1\}$ remains unexposed. This method does not expose any additional endpoints, and in fact the identity of the endpoint $\mathbf{a}_0$ (as well as the vertex $\mathbf{u}_1$) remains unknown after this procedure. The state after the structural modifications is shown in figure 11.2, and since all exposed edges are removed, it follows that conditional uniformity is maintained.

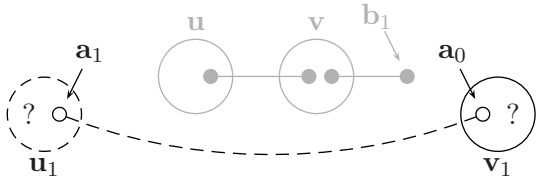


Figure 11.2: The state of the modified greedy heuristic after structural modifications; the greyed out vertices and endpoints are removed, and the endpoint $\mathbf{a}_0$, which previously belonged to the vertex $\mathbf{u}$, is re-assigned to the vertex $\mathbf{v}_1$. Note that the edge connecting $\mathbf{a}_0$ and $\mathbf{a}_1$ is still unexposed, and the identity of the vertex $\mathbf{u}_1$ is still unknown.

□

Hence, this heuristic preserves conditional uniformity, and we now proceed to compute the expected increments of the 4 parameters $\mathbf{m}_t$, $\mathbf{n}_t$, $\mathbf{q}_t$, and $\mathbf{\Lambda}_t$.

Lemma 11.3.4. *The modified greedy heuristic achieves expected increments of*

$$\begin{aligned}
\mathsf{E}_t[\Delta\mathbf{m}_{t+1}] &= -f^*(\mathbf{z}_t) & f^*(z) &= 4 + 8k_1(z) = 4 + 8(3-6z)^2 \\
\mathsf{E}_t[\Delta\mathbf{n}_{t+1}] &= -g^*(\mathbf{z}_t) & g^*(z) &= 2 - k_1(z) = 2 + (3-6z)^2 \\
\mathsf{E}_t[\Delta\mathbf{q}_{t+1}] &= h^*(\mathbf{z}_t) & h^*(z) &= 1 \\
\mathsf{E}_t[\Delta\mathbf{\Lambda}_{t+1}] &= k^*(\mathbf{z}_t) & k^*(z) &= 4k_1(z)(6z-2) = 4(3-6z)^2(6z-2),
\end{aligned}$$

whenever the minimum degree is 2*, and within an additive factor of* $\pm o(1)$*.*

Proof. For simplicity, we ignore the additive factor of $\pm o(1)$ in our computations which arises from various approximations and minor dependencies; in particular, probability of

encountering a loop or a set of parallel edges during any given iteration is $o(1)$ we may ignore this possibility and still achieve the desired accuracy of approximation.

Now, the probability that a randomly selected endpoint belongs to a vertex of degree 3 (i.e. has residual degree 2) is equal to $\mu(2) = (3-6z)$. The vertex v chosen by the heuristic has two neighbors, and hence the probability that both have degree 3 is

$$\mu(2)^2 = (3-6z)^2 = (3-6z)^2 = k_1(z).$$

In this case, as with the greedy heuristic, we remove v along with both of its neighbors, for a total of 3 vertices. And, since each neighbor has 2 additional incident edges, a total of 6 edges and therefore 12 endpoints are removed.

Otherwise, with probability $1 - \mu(2)^2$, at least one neighbor has degree 2, in which case we we simply remove 2 vertices and 4 endpoints as shown above. It follows that the expected number of vertices removed is

$$g^*(z) = 3(k_1(z) + 2(1 - k_1(z) = 2 + k_1(z),$$

and the expected number of endpoints removed is

$$f^*(z) = 12k_1(z) + 4(1 - k_1(z)) = 4 + 8k_1(z).$$

And, clearly, unless a loop occurs, exactly one vertex is added to the independent set, so $h^*(z) = 1$.

Finally, if at least one neighbor has degree 2, than all of the vertices which remain in the graph retain their original degree, so in this case no new vertices of degree 1 are created. Otherwise, if both neighbors have degree 3, then 4 additional endpoints are removed uniformly at random, corresponding to the 4 additional edges incident on these two neighbors. Each one of these creates a new vertex of degree 1 with probability $\mu(1) = 6z - 2$, and therefore the expected number of such vertices is

$$k^*(z) = 4k_1(z)(6z - 2),$$

□

As with the greedy heuristic analyzed above, the remaining portion of the proof of Theorem 11.3.2 involves routine computations.

Proof of Theorem 11.2.1. We first compute the fraction

$$j(z) = k^*(z)/k_1(z) = \frac{4k_1(z)(6z-2)}{k_1(z)} = 4(6z-2).$$

Next, we examine the adjusted expected increments $f(z)$, $g(z)$, and $h(z)$ from equation (11.8) proposition 11.1.5. In this case, the function $f(z) = f^*(z) + f_1(z)j(z)$ has a simple expression:

$$\begin{aligned} f(z) = f^*(z) + f_1(z)j(z) &= 4(1 + 2k_1(z) + 2(6z-2)(5-6z)) \\ &= 4(1 + 2(3-6z)^2 + 2(6z-2)(5-6z)) \\ &= 4(12z-1). \end{aligned} \tag{11.9}$$

Since

$$\begin{aligned} g(z) = g^*(z) + g_1(z)j(z) &= k_1(z) + 2 + 8(6z-2) \\ &= (3-6z)^2 + 2 + 8(6z-2), \end{aligned}$$

then we may also obtain a reasonably simple expression for $z - g(z)/f(z)$:

$$\begin{aligned} z - g(z)/f(z) &= z - \frac{(3-6z)^2 + 2 + 8(6z-2)}{4(12z-1)} \\ &= \frac{4z(12z-1) - (3-6z)^2 - (48z-14)}{4(12z-1)} \\ &= \frac{12z^2 - 16z + 5}{4(12z-1)} \\ &= \frac{(1-2z)(5-6z)}{4(12z-1)}. \end{aligned} \tag{11.10}$$

Therefore

$$\frac{1}{z - g(z)/f(z)} = \frac{4(12z-1)}{(1-2z)(5-6z)} = \frac{10}{1-2z} - \frac{54}{5-6z},$$

which is easily integrated:

$$r(z) = \int_{y=1/3}^{z} \frac{dy}{y - g(y)/f(y)} = 9\ln(5-6z) - 5\ln(1-2z) - 14\ln 3.$$

This yields the following expression for $m(z)$:

$$\begin{aligned} m(z) = 3e^{-r(z)} &= 3^{15}(5-6z)^{-9}(1-2z)^5 \\ &= \frac{3^{10}(3-6z)^5}{(5-6z)^9}. \end{aligned} \tag{11.11}$$

Again, for a "sanity check," we confirm that $m(1/3) = 3^{10}3^{-9}1^5 = 3$ and $m(1/2) = 3^{10}2^{-9}0^5 = 0$.

Next, we compute

$$\begin{aligned} h(z) = h^*(z) + h_1(z)j(z) &= 1 + 4(6z-2) \\ &= 24z - 7, \end{aligned}$$

which combined with (11.9) and (11.10) yields

$$\begin{aligned} \frac{h(z)/f(z)}{z - g(z)/f(z)} &= \left(\frac{24z-7}{4(12z-1)}\right)\left(\frac{4(12z-1)}{(1-2z)(5-6z)}\right) \\ &= \frac{24z-7}{(1-2z)(5-6z)}. \end{aligned}$$

By Theorem (11.1.1), the lower bound on γ is achieved by integrating $\frac{m(z)h(z)/f(z)}{z-g(z)/f(z)}$ from $z = 1/3$ to $z = 1/2$, and using the above equation, along with the expression (11.11) for $m(z)$, we have

$$\frac{m(z)h(z)/f(z)}{z - g(z)/f(z)} = \frac{3^{11}(24z-7)(3-6z)^4}{(z-6z)^{10}}, . \tag{11.12}$$

In order to integrate the right hand side of (11.12), we change variables and let

$y = (5 - 6z)$:

$$\begin{aligned}
\int_{z=1/3}^{1/2} \frac{3^{11}(24z-7)(3-6z)^4}{(5-6z)^{10}} dz &= \int_{y=3}^{2} \frac{3^{11}(13-4y)(y-2)^4}{y^{10}}(-dy/6) \\
&= (3^{10}/2)\int_{y=2}^{3} \frac{3^{11}(13-4y)(y-2)^4}{y^{10}} dy \\
&= (3^{10}/2)\int_{y=2}^{3} \frac{5(y-2)^4 - 4(y-2)^5}{y^{10}} dy \\
&= (3^{10}/2)\int_{y=2}^{3} \left(5\sum_{i=0}^{4} \binom{4}{i} y^{i-10}(-2)^{5-i} - 4\sum_{j=0}^{5} \binom{5}{j} y^{j-10}(-2)^{4-j} \right) dy \\
&= (3^{10}/2)\int_{y=2}^{3} \sum_{i=0}^{5} ((5-i) - 4(-2)) \binom{5}{i} y^{i-10}(-2)^{4-i} dy \\
&= (3^{10}2^3)\int_{y=2}^{3} \sum_{i=0}^{5} (13-i) \binom{5}{i} y^{i-10}(-2)^{-i} dy.
\end{aligned}$$

Computing a solution is now a manner of ordinary arithmetic, and since the limits of integration are integers, then the solution is in fact a rational number:

$$\begin{aligned}
\gamma &\geq (3^{10}2^3)\int_{y=2}^{3} \sum_{i=0}^{5} (13-i) \binom{5}{i} y^{i-10}(-2)^{-i} \\
&= (3^{10}2^3)\sum_{i=0}^{5} \binom{5}{i} (-2)^{-i}(13-i)\left(\frac{2^{i-9} - 3^{i-9}}{9-i} \right) \\
&= \frac{1181}{2688}.
\end{aligned}$$

In decimal form, this yields a bound of $\gamma \geq 0.43936\ldots$ □

11.4 Generalizations and Open Questions

Due to the ease of solving the systems differential equations involved, random 3-regular graphs provide an opportunity to push the limits of the algorithmic method without getting bogged down by infinite systems of differential equations. The technical problems related to convergence etc. are all relatively trivial, and thus we are free to explore the limits of the algorithmic method. As such, we now discuss various possible generalizations and

open problems, many of which would prove quite difficult in the general setting of the configuration model, but are perhaps more approachable for 3-regular graphs.

11.4.1 Marginal Improvements

It is fairly evident that the modified greedy heuristic introduced in section 11.3 is not optimal, and in fact it does not appear difficult to improve upon the lower bound computed above using the same basic techniques.

Specifically, recall that the modified greedy heuristic behaves optimally unless the vertex v of degree 2 has two neighbors of degree 3. Now, in this situation, any maximal independent set must contain either v or one of its neighbors. And, it is easy to see that there exists a maximal independent set which contains either v or both neighbors.

Without additional information, it appears preferable to add v to the independent set, rather than one or both of the neighbors. Intuitively, this is because adding v to the independent set achieves a ratio of 1 independent vertex per 3 removed vertices, while adding both neighbors results in 2 independent set vertices and a total of 7 vertices removed, for a ratio of $2/7 < 1/3$.

Of course, as noted in section 11.4, the optimal heuristic behavior does not depend only on the ratio independent set vertices to total vertices removed. In order to truly decide which behavior is preferable, we must revisit the inequality in (11.6):

$$z^2\gamma'(z) \leq \frac{\gamma(z) - h(z)/g(z)}{f(z)/g(z) - 1/z}.$$

In addition to the ratio $h(z)/g(z)$ of independent set vertices to vertices removed, this expression also depends on the ratio $f(z)/g(z)$ of endpoints to vertices removed, as well as the optimal independence ratio $\gamma(z)$ given the current value of $z = n/m$. Moreover, even this expression does not accurately determine which choice is preferable, since it assumes that no additional vertices of degree 1 are created by the heuristic. In order to make the best possible decision given the current state information, we must also take into account the fact that each time we remove an endpoint uniformly at random, a new vertex of degree 1 is created with probability $\mu(1) = 6z - 2$, and then adjust the above expression accordingly.

We omit the proof in this informally discussion, but it can be shown that, given only the fact that v has degree 2 and both neighbors have degree 3, it is indeed preferable to add the vertex v to the independent set rather than both neighbors. However, if we were to expose a slightly larger neighborhood of v before making this decision, the situation might change. For example, suppose that all 4 of the second generation neighbors of the vertex v also have degree 3; this situation is illustrated in figure 11.3. In this case, for a number of reasons, it is less apparent that adding v to the independent set is the best choice, and the optimal behavior will most likely depend on the value of z and $\gamma(z)$ in the above equation.

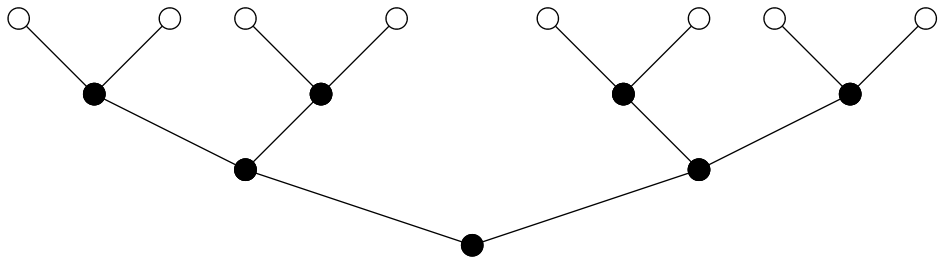


Figure 11.3: In this situation, it might be unwise to add v to the independent set, despite the fact that both neighbors have degree 3.

And, as we continue to explore larger neighborhoods of v, the additional information we receive will allow us to make progressively better decisions about which vertices should be added to the independent set. We can thus imagine a class of modified greedy heuristics which behave as follows:

1. choose a vertex v and, for a fixed constant C, expose all edges within distance C of v;
2. decide which vertices in this neighborhood belong in the independent set and which do not;
3. remove these vertices and their neighbors, and proceed recursively.

Now, for any given constant C, it appears straightforward, though perhaps not entirely trivial, to perform the necessary computations to decide how best to assign vertices to the

independent set in the second step of the heuristic described above. Indeed, since there are only finitely many possible local structures, we can proceed by "brute force" if necessary, by simply trying every possible combination and cranking through the formula from Theorem 11.1.1 to see which yields the best result. As noted above, the best decision may depend on both z and $\gamma(z)$, but, even so, the problem of finding the best best lower bound on γ for the above heuristic is not particularly difficult, assuming that we are willing to put in the necessary effort.

Let us now define by γ_C the best lower bound on the independence ratio γ of 3-regular graphs which can be computed by a modified greedy heuristic which explores a neighborhood of depth C as described above. It is reasonable to expect that increasing the depth of the local search will yield a better lower bound, and if this is the case we will have a sequence $\gamma_1 < \gamma_2 < \gamma_3 < \ldots \gamma$ of lower bounds to γ. And, for any fixed C, the lower bound γ_C is not difficult to compute, at least in principle.

However, it is not particularly interesting to simply compute γ_C for progressively larger values of C, since if the above intuition is correct, this will simply lead to a sequence of better lower bounds, each of which requires slightly more work to compute than the last. We thus propose the following open problem.

- **Open Problem 1**: Compute the limit

$$\gamma^* = \lim_{C \to \infty} \gamma_C$$

of the lower bounds $\gamma_C < \gamma$ achievable by modified greedy heuristics as described above.

It does not appear impossibly difficult to compute this limit, but, unlike the values γ_C, the limit γ^* cannot be determined by brute force computation alone.

Now, for various reasons which are beyond the scope of this discussion, it is extremely unlikely that the limit γ^* will be equal to the true independence ratio γ of random 3-regular graphs. But, unlike any particular lower bound γ_C for any fixed C, the limit γ^* might be considered a "natural" threshold of some sort. Hence, we offer a second open problem.

- **Open Problem 2**: Does the limit γ^* carry any structural significance?

This problem is of course open-ended, since "structural significance" not well-defined; a more precise definition of this notion is possible but beyond the scope of this discussion.

11.4.2 Other Optimization Problems

There are typically two stages involved in the algorithmic analysis of any random graph property: first, we must devise an algorithm which computes the given graph property, and second, we must analyze the execution of this algorithm on a random input graph. And, typically, analyzing the execution is the more difficult of the two tasks, since it involves solving differential equations and so on.

However, for the special case of i-set heuristics on random 3-regular graphs, the parameter space is essentially one-dimensional, and we have been able to exploit this fact in the form of Theorem 11.1.1, which gives a general solution to the execution of any such heuristic based on expected increments. As a result, the second, and generally more difficult, portion of the algorithmic analysis becomes relatively "easy," and we are free to dedicate more effort to creating stronger heuristics.

Theorem 11.1.1 simply represents a general solution to a certain (simple) class of linear differential equations, translated into a form which is applicable to the analysis of heuristics on random 3-graphs, and there is nothing about this solution which is uniquely applicable to the independent set problem. Specifically, in order to reduce the parameter space to a single direction, all that is necessary is that the number of vertices of degree 1 remains $o(n)$ throughout the algorithm. For the independent set problem, this will always be the case, since whenever one or more vertices of degree 1 are present, we may simply greedily add such vertices to our independent set.

In general, though, the same methodology will be applicable to other optimization problems, provided that there is some canonical method to dispose of vertices of degree 1. In this case, we may the parameter $\mathbf{q}_t$ may represent some other quantity we wish to optimize, and the formula computed in 11.1.1 will hold in precisely the same way. Of course,

the exact functions f_1, g_1, h_1, and k_1 from proposition 11.1.4, as well as the formula from proposition 11.1.5 which implicitly computes the relative frequency of iterations of degrees 1 and 2 will depend the procedure used when vertices of degree 1 are present. But, overall, the methodology developed in section 11.4 can be applied to other optimization problems with relatively minor modifications.

One particular problem which might be successfully analyzed using these techniques is the *dominating set* problem. Recall that a dominating set in a graph is a vertex subset $W \subseteq V$ such that every $v \in V$ is adjacent to at least on $w \in W$. Finding a dominating set of minimum size is a well-known NP-hard optimization problem.

And, as with the independent set problem, there is a simple and optimal method of dealing with vertices of degree 1: if v has degree 1, then we may simply add its neighbor u to the dominating set, and then remove v and u, as well as all additional neighbors of the vertex u. As noted above, the expected increments of this operation will not be quite the same as for the independent set heuristic, but the basic technique from section 11.4 should be applicable with minor modifications.

Appendix

Appendix A

Simulating $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$ with the Configuration Model

The classical Erdős-Rényi random graphs $\mathbf{G}_{n,m}$ and $\mathbf{G}_{n,p}$ are models for random simple graphs on n vertices:

- $\mathbf{G}_{n,m}$ is chosen uniformly at random from all simple graphs with m edges and n vertices;
- in $\mathbf{G}_{n,p}$ each possible edge $\{u, v\}$ occurs independently with probability p.

We show how the Erdős-Rényi random graphs can be simulated using the configuration model.

First, recall that the "random graph" $\boldsymbol{\mathcal{G}}(H)$ is in fact a random configuration $(A, V, \mathbf{E})$. As described in chapter 3, we may obtain an ordinary graph $\hat{G} = (V, \hat{E})$ from any configuration $G = (A, V, E)$ by associating a graph edge $\hat{e} = (V(a_1), V(a_2))$ with every configuration edge $e = \{a_1, a_2\}$. Accordingly, the configuration model can be used to generate random ordinary graphs $\hat{\boldsymbol{\mathcal{G}}}(H)$, and by specifying a Poisson residual distribution, we may simulate the Erdős-Rényi random graphs as follows.

Theorem A.0.1. *For fixed $c > 0$, let $\boldsymbol{\mathcal{G}}(H) = \boldsymbol{\mathcal{G}}(\pi_c)$ be an asymptotic random configuration with Poisson residual distribution*

$$\pi_c(i) = \frac{e^{-c}c^i}{i!},$$

and also assume that the fraction $\lambda_H(0)$ of vertices of degree 0 converges to $\pi_c(0) = e^{-c}$.

1. *Any property which is satisfied w.e.h.p. by $\hat{\boldsymbol{\mathcal{G}}}(H)$ will also by satisfied w.e.h.p. by the Erdős-Rényi random graph $\mathbf{G}_{n,m}$ with $m/n \to c/2$.*
2. *If we are impose the additional assumptions:*

(a) the maximum degree is at most $\epsilon \ln n$ for fixed $\epsilon > 0$,

(b) the first moment $M_k(\mu_H)$ converges to $M_k(\pi_c) = c^k$,

then any property which is satisfied w.p.h.p. by $\hat{\boldsymbol{\mathcal{G}}}(H)$ will also hold w.p.h.p. for $\mathbf{G}_{n,m}$ with $m/n \to c/2$.

3. *In both cases, the high-probability guarantee can be weakened; for example, any property which is satisfied a.a.s. by $\boldsymbol{\mathcal{G}}(H)$ will also hold a.a.s. for $\mathbf{G}_{n,m}$.*

While this theorem nominally deals only with $\mathbf{G}_{n,m}$, it is well-known (e.g. [13]) that $\mathbf{G}_{n,p}$ and $\mathbf{G}_{n,m}$ are essentially equivalent; indeed, $\mathbf{G}_{n,p}$ is distributed identically to $\mathbf{G}_{n,m}$ conditional on the number of edges, and the number of edges is the sum of independent Bernoulli trials and is thus w.e.h.p. concentrated. When working with the configuration model, though it is simpler to simulate $\mathbf{G}_{n,m}$ rather than $\mathbf{G}_{n,p}$ due the the fact that the number of edges is known.

For simplicity, we shall work with ordinary graphs on the canonical n-vertex set $V = [n] = \{1, \ldots, n\}$, and we recall the following definitions:

- a *simple graph* is pair $(V = [n], \hat{E})$, where $\hat{E}$ is a set of unordered pairs of distinct vertices;
- a *multi-graph* is a pair $(V = [n], \hat{E})$ where $\hat{E}$ is a multi-set of unordered pairs of not necessarily distinct vertices.

Since there are $\binom{n}{2}$ possible simple edges, then the number of simple graphs with m edges is $\binom{\binom{n}{2}}{m}$.

A.1 Endpoint Partitions, Configurations, and Ordinary Graphs

The endpoint configurations and configurations we work with will also be restricted to a canonical form, in which the endpoint set is of the form $A = [m] = \{1, \ldots, m\}$. Hence, we shall let $\mathbb{H}(n, m)$ denote the set of all endpoint partitions with vertex set $[n]$ and endpoint

set $[m]$. Any element of $\mathbb{H}(n,m)$ can thus be characterized by the mapping $V : [m] \to [n]$ assigning endpoints to vertices. In particular, there are m^n such endpoint partitions, and we denote a uniformly random element of $\mathbb{H}(n,m)$ by $\boldsymbol{\mathcal{H}}(n,m)$.

Accordingly, for even values of m, we let $\mathbb{G}(n,m) = \mathbb{G}(\mathbb{H}(n,m))$ denote the set of configurations with canonical endpoint partitions. The number of such configurations is easily seen to me

$$|\mathbb{G}(n,m)| = |\mathbb{H}(n,m)| \cdot (m-1)!! = m^n(m-1)!!.$$

Of course many of these will be isomorphic, but since there are many equivalence relations, we use the set $\mathbb{G}(n,m)$ as a starting point.

We let

$$\boldsymbol{\mathcal{G}}(n,m) = \boldsymbol{\mathcal{G}}(\boldsymbol{\mathcal{H}}(n,m))$$

denote the corresponding random configuration, noting that there is slight inconsistency in this notation, since $\mathbf{G}_{n,m}$ is a random ordinary graph with exactly m edges, while $\boldsymbol{\mathcal{G}}(n,m)$ contains m endpoints and therefore $m/2$ edges. Hence, our objective is to model the Erdős-Rényi random graph $\mathbf{G}_{n,m}$ using the simple graph $\hat{\boldsymbol{\mathcal{G}}}(n,2m)$ corresponding to the random configuration $\boldsymbol{\mathcal{G}}(n,2m)$.

We begin by noting that $\boldsymbol{\mathcal{G}}(n,2m)$ can be generated as follows.

Proposition A.1.1. *Let $(\mathbf{v}_1,\dots,\mathbf{v}_{2m})$ be a sequence of i.i.d. uniformly random elements of $[n]$ (without replacement). Then the set of graph-edges in the random ordinary graph $\hat{\boldsymbol{\mathcal{G}}}(n,2m)$ is distributed identically to*

$$\mathcal{E}(\mathbf{v}_1,\dots,\mathbf{v}_{2m}) = \{\{\mathbf{v}_1,\mathbf{v}_2\},\dots,\{\mathbf{v}_{2m-1},\mathbf{v}_{2m}\}. \tag{A.1}$$

Proof. Choose a uniformly random permutation $(\mathbf{a}_1,\dots,\mathbf{a}_{2m})$ of $[m]$, and let $(\mathbf{v}_1,\dots,\mathbf{v}_{2m})$ denote a sequence of $2m$ i.i.d. uniformly random elements from $[n]$ as described above. We claim that $\boldsymbol{\mathcal{G}}(n,2m)$ can be generated by letting $\mathbf{V}(\mathbf{a}_t) = \mathbf{v}_t$ for each t, and using the matching

$$\mathbf{E} = \mathcal{E}(\mathbf{a}_1,\dots,\mathbf{a}_{2m}) = \{\{\mathbf{a}_1,\mathbf{a}_2\},\dots,\{\mathbf{a}_{2m-1},\mathbf{a}_{2m}\}\}$$

as the set of configuration-edges.

Clearly, the edge set is a uniformly random matching of $[2m]$. Also, the vertex assignment $\mathbf{V} : [2m] \to [n]$ is uniformly random, and this assignment is independent of the edge set, and therefore this configuration is distributed identically to $\boldsymbol{\mathcal{G}}(n, 2m)$.

The set of graph-edges in $\boldsymbol{\mathcal{G}}(n, 2m)$ in this case is given by

$$\hat{\mathbf{E}} = \{\{\mathbf{v}_1, \mathbf{v}_2\}, \ldots, \{\mathbf{v}_{2m-1}, \mathbf{v}_{2m}\}\},$$

and the proof is complete. □

Corollary A.1.2. *Every simple graph with m edges and n vertices occurs with equal probability as $\hat{\boldsymbol{\mathcal{G}}}(n, 2m)$, and if both $n \geq 4$ and $m \leq n^2/8$, then the probability that $\hat{\boldsymbol{\mathcal{G}}}(n, 2m)$ is simple is at least*

$$e^{-2m(n+2m)/n^2}$$

Proof. Since the set of graph-edges in a simple graph contains no loops or parallel edges, it is clear that there are precisely $2^m m!$ sequences $(v_1, \ldots, v_{2m})$ of vertices that correspond to each simple graph, and therefore each simple graph occurs with equal probability.

For the bound on the probability that m is simple, for each $r \leq m$, let L_r denote the event that no loops or parallel edges occur in the partial edge set $\{\{\mathbf{v}_1, \mathbf{v}_2\}, \ldots, \{\mathbf{v}_{2r-1}, \mathbf{v}_{2r}\}\}$. Since there are $\binom{n}{2}$ possible non-loop edges, then, conditional on L_r, there are $\binom{n}{2} - r$ non-loop edges which will not create a parallel edge, and hence we may compute

$$\mathsf{P}[L_{r+1} \mid L_r] = \frac{n(n-1) - 2r}{n^2} \geq 1 - \frac{n + 2r}{n^2} \geq 1 - \frac{n + 2m}{n^2}.$$

If $m \leq n^2/4$ and $n \geq 4$

$$n + 2m \leq n + n^2/4 \leq n^2/2,$$

and therefore $\frac{n+2m}{n^2} < 1/2$, and we may use the approximation

$$1 - \frac{n + 2m}{n^2} > e^{-2\left(\frac{n+2m}{n^2}\right)}.$$

We now conclude that the probability that $\hat{\boldsymbol{\mathcal{G}}}(n, 2m)$ is simple is at least

$$\begin{aligned}\mathsf{P}[L_m] = \prod_{r=1}^{m} \mathsf{P}[L_r \mid L_{r-1}] &\geq \prod_{r=1}^{m} \left(1 - \frac{n+2m}{n^2}\right) \\ &\geq e^{-2m(n+2m)/n^2}.\end{aligned}$$

□

Corollary A.1.3. *For any fixed $c > 0$, if $m/n \to c/2$, any property which is satisfied w.e.h.p., w.p.h.p., or a.a.s. by $\hat{\boldsymbol{\mathcal{G}}}(n, 2m)$ is also satisfied, with the same high-probability guarantee, by the Erdős-Rényi random graph $\mathbf{G}_{n,m}$.*

Proof. Choose any property α, and note that, since every simple graph with n vertices and m edges occurs with equal probability as $\hat{\boldsymbol{\mathcal{G}}}(n, 2m)$ then

$$\mathsf{P}[\alpha(\mathbf{G}_{n,m}) = \top] = \mathsf{P}[\alpha(\hat{\boldsymbol{\mathcal{G}}}(n, 2m) = \top \mid \hat{\boldsymbol{\mathcal{G}}}(n, 2m) \text{ is simple}].$$

The probability that $\hat{\boldsymbol{\mathcal{G}}}(n, 2m)$ is simple under these assumptions is $e^{-2m(n+2m)/n^2} = e^{-O(1)} = \Omega(1)$, and it follows that

$$1 - \mathsf{P}[\alpha(\mathbf{G}_{n,m}) = \top] = O\Big(1 - \mathsf{P}[\alpha(\hat{\boldsymbol{\mathcal{G}}}(n, 2m) = \top]\Big),$$

so any high-probability guarantee which holds for $\hat{\boldsymbol{\mathcal{G}}}(n, 2m)$ will hold for for $\mathbf{G}_{n,m}$ as well. □

A.2 Poisson Convergence

Proposition A.2.1. *Let $\boldsymbol{\lambda}(n, m)$ denote the degree distribution of the random endpoint partition $\boldsymbol{\mathcal{H}}(n, m)$. Then, for any fixed $c \geq 0$, if $m/n \to c$, then $\boldsymbol{\lambda}(n, m)$ converges w.e.h.p. (with respect to n) to the Poisson distribution π_c in the limit as $n \to \infty$.*

Proof. We may use the differential equations method to establish this w.e.h.p. convergence as follows. Let $\boldsymbol{\lambda}_t$ denote the degree distribution after m random endpoints have been added, and note that clearly $\Delta\boldsymbol{\lambda}_t(i) \leq \frac{1}{n}$, since at most one vertex changes degree each step. Also we have

$$\mathsf{E}_t[\Delta\boldsymbol{\lambda}_{t+1}(i)] = \frac{\boldsymbol{\lambda}_t(i-1) - \boldsymbol{\lambda}_t(i)}{n},$$

since $\boldsymbol{\lambda}_{t+1}$ increases or decreases by $1/n$ if the endpoint $\mathbf{a}_{t+1}$ is assigned to a vertex of degree $i-1$ or i, respectively.

We scale by dividing by n, so $\boldsymbol{\lambda}_\xi = \boldsymbol{\lambda}_{\lfloor \xi n \rfloor}$, and thus by martingale concentration and Doob decomposition, we have the system of differential equations

$$d\lambda_\xi(i) = (\lambda_\xi(i-1) - \lambda_\xi(i))d\xi$$

which must be satisfied by every w.e.h.p. solution

In this case, there is no possibility of a non-analytic solution, since each finite system of $\lambda_\xi(i)$ for $i \in \{0, \ldots, J\}$ has a unique solution given by

$$\lambda_\xi(i) = \frac{e^{-\xi}\xi^i}{i!} = \pi_c(i),$$

which can be verified by computing

$$d\lambda_\xi(i) = \frac{ie^{-\xi}\xi^{i-1} - e^{-\xi}\xi^i}{i!} = \lambda_\xi(i-1) - \lambda_\xi(i).$$

Since the process terminates at time $t = m = c/n$, the unique w.e.h.p. solution to the degree distribution at this time is $\lambda_\xi = \pi_c$. □

Corollary A.2.2. *If $m/n \to c$, then the residual distribution $\boldsymbol{\mu}(n,m)$ of $\boldsymbol{\mathcal{H}}(n,m)$ converges w.e.h.p. to π_c as well.*

Proof. Clearly, the average vertex degree $M(\lambda(n,m)) = m/n \to c$, and therefore we may compute

$$\boldsymbol{\mu}(i) = \frac{(i+1)\boldsymbol{\lambda}(i+1)}{c} \to \frac{(i+1)e^{-c}c^{(i+1)}}{(i+1)!}\frac{e^{-c}c^i}{i!} == \pi_c(i).$$

□

A.3 Maximum Degree and the First Moment

Proposition A.3.1. *For any $\epsilon > 0$, the maximum degree in $\boldsymbol{\mathcal{H}}(n,m)$ with $m/n \to c$ at most $\epsilon \ln n$ w.p.h.p.*

Proof. It is clear that the degree of any given vertex in $\boldsymbol{\mathcal{G}}(n,m)$ is binomially distributed, and we may thus compute

$$\mathsf{P}[\mathbf{deg}(v) \geq d] \leq \binom{m}{d}(1/n)^d \leq (m/n)^d/d!.$$

If $d = \epsilon \ln n$, then this probability is $n^{-\omega(1)}$, and the probability that this occurs for any of the n vertices is also $n^{-\omega(1)}$. □

Since $M(\mu) = \frac{M_2(\lambda)}{M(\lambda)}$, in order to establish convergence of the first moment of the residual distribution, it suffices to show convergence of the second moment of the degree distribution.

Proposition A.3.2. *If $m/n \to c$, then the second moment $M_2(\boldsymbol{\lambda}(n,m))$ converges w.p.h.p. to c^2.*

Proof. Again we consider the process $\boldsymbol{\lambda}_t$ in which one endpoint is added each step, and we let $\boldsymbol{\tau}$ denote the first time that the maximum degree exceeds $\ln n$. By the above proposition, this never occurs and therefore $\boldsymbol{\tau} = m$ w.p.h.p.

Also, note that the increments of $M_2(\boldsymbol{\lambda}_t)$ are given by

$$\Delta M_2(\boldsymbol{\lambda}_{t+1}) = \frac{\sum_{v \in V} \Delta\Big(\mathbf{deg}_{t+1}(v)(\mathbf{deg}_{t+1}(v)-1)\Big)}{n},$$

Since only one vertex $\mathbf{v}_t$ changes degree each step, then

$$\begin{aligned}\Delta M_2(\boldsymbol{\lambda}_{t+1}) &= \frac{(\mathbf{deg}_t(\mathbf{v}_t)+1)\,\mathbf{deg}_t(\mathbf{v}) - \mathbf{deg}_t(\mathbf{v}_t)(\mathbf{deg}_t(\mathbf{v}_t)-1)}{n} \\ &= \frac{2\,\mathbf{deg}_t(v_t)}{n},\end{aligned}$$

and since $\mathbf{v}_t$ is chosen uniformly at random, it follows that

$$\mathsf{E}_t[\Delta M_2(\boldsymbol{\lambda}_{t+1})] = 2M(\boldsymbol{\lambda}_t) = 2t/n.$$

Since we have stopped the process when the maximum degree exceeds $\ln n$, the the increments of this process are bounded by $\ln n/n$, and hence Azuma's inequality implies

that the terminal value concentrates around it expectation w.p.h.p., and hence we have the w.p.h.p. differential constraint

$$dM_2(\lambda_\xi) = 2\xi d\xi,$$

which has solution $M_2(\lambda_\xi) = \xi^2$, and therefore if $m/n = c$, the second factorial moment converges w.p.h.p. to c^2. □

The proof of Theorem A.0.1 is essentially a summary of the above propositions. Before we give the proof, we point out that the maximum degree bound and the convergence of the first moment of the residual distribution both hold with stronger than polynomially high probability, but w.p.h.p. condition is all we need in this dissertation. Also, higher moments of the residual distribution can be shown to converge using essentially the same argument, but again this is unnecessary for our purposes.

Proof of Theorem A.0.1. By corollary A.1.3, high-probability guarantees for $\mathbf{G}_{n,m}$ (and therefore $\mathbf{G}_{n,p}$) can be achieved using the random configuration $\boldsymbol{\mathcal{G}}(n, 2m)$, provided that $m/n \to c/2$. In this case, the degree and residual distribution of $\boldsymbol{\mathcal{G}}(n, 2m)$ will both converge w.e.h.p. to the Poisson distribution. Also, the bound on the maximum degree and the convergence of first moment $M(\boldsymbol{\mu})$ hold w.p.h.p. Hence, of $\boldsymbol{\mathcal{G}}(\pi_c)$ can be shown to satisfy a given property either w.e.h.p., or w.p.h.p. with the additional assumptions of the degree bound and convergence of the first moment, then the same will hold for the Erdős-Rényi random graph $\mathbf{G}_{n,m}$. □

Bibliography

[1] D. Achlioptas and C. Moore. Almost all graphs with degree 4 are 3-colorable. *JCSS*, 67(2):441–471, 2003.

[2] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *Proc. STOC*, pages 171–180, 2000.

[3] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley and Sons, 2000.

[4] K. B. Athreya and P. E. Ney. *Branching Processes*. Springer-Verlag, 2000.

[5] Y. Azar, A. Z. Broder, A. R. Karlin, and E. Upfal. Balanced allocations. In *Proc. STOC*, pages 593–602, 1994.

[6] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku. Math. J.*, 3:357–367, 1967.

[7] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[8] A. Barabási and R. Albert. Mean-field theory for scale-free random networks. *Physica A*, 272:173–187, 1999.

[9] E. A. Bender and E. R. Canfield. The asymptotic number of labelled graphs with given degree sequences. *Journal of Combinatorial Theory (A)*, 24:296–307, 1978.

[10] P. Billingsley. *Probability and Measure*. John Wiley & Sons, 1995.

[11] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J Combinatorics*, 1:311–316, 1980.

[12] B. Bollobás. The independence ratio of regular graphs. *Proc. Amer. Math. Soc.*, 83:433–436, 1981.

[13] B. Bollobás. *Random Graphs*. Academic Press, 2000.

[14] B. Bollobás. *Modern Graph Theory*. Springer, 2002.

[15] B. Bollobás and W. F. de la Vega. *Combinatorica*, 2:125–134, 1982.

[16] A. Braunstein, M. Mézard, and R. Zecchina. Survey propagation: an algorithm for satisfiability. *Random Structures and Algorithms*, 27:201–226, 2005.

[17] S. N. Burris and H. P. Sankappanavar. *A Course in Universal Algebra*. Springer-Verlag, 1981. `http://www.math.uwaterloo.ca/ snburris/htdocs/ualg.html`.

[18] J. Cain, P. Sanders, and N. Wormald. The random graph threshold for k-orientability. In *Proc. SODA*, 2007.

[19] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23:493–507, 1952.

[20] F. Chung and L. Lu. The diameter of random sparse graphs. *Advances in Applied Math.*, 26:257–279, 2001.

[21] W. Coddington. *An Introduction to Ordinary Differential Equations*. Dover, 1961.

[22] C. Cooper. The cores of random hypergraphs with a given degree sequence. *Random Structures and Algorithms*, 25:353–375, 2004.

[23] A. Czumaj and V. Stemmann. Randomized allocation processes. *RSA*, 18:297–331, 2001.

[24] D. J. de Solla Price. Networks of scientific papers. *Science*, 149(3683):510–515, 1965.

[25] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, 1998.

[26] M. Dietzfelbinger and P. Wielding. Balanced allocation and dictionaries with packed constant size binds. In *Proc. ICALP*, 2005.

[27] P. Erdős and A. Rényi. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc*, 53:292–294, 1947.

[28] P. Erdős and A. Rényi. On random graphs. *Publ. Math. Debrecen*, 6:290–297, 1959.

[29] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Int. Statis. Tokyo*, 38:343–347, 1961.

[30] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM*, pages 251–262, 1999.

[31] D. Fernholz. k-orientability, probability generating functions, and algorithms on random graphs. 2006. Manuscript.

[32] D. Fernholz and V. Ramachandran. Cores and connectivity in sparse random graphs. Technical Report UTCS TR04-13, Dept of Comp Sci, Univ Texas, Austin, 2004.

[33] D. Fernholz and V. Ramachandran. The diameter of sparse random graphs. Technical Report UTCS TR04-53, Department of Computer Sciences, University of Texas, Austin TX 78712, 2004. Submitted to *Random Structures an Algorithms*.

[34] D. Fernholz and V. Ramachandran. The k-orientability thresholds for $G_{n,p}$. In *SODA*, 2007.

[35] A. Frieze and C. McDiarmid. Algorithmic theory of sparse random graphs. *Random Structures Algorithms*, 10:5–42, 1997.

[36] A. Frieze and S. Suen. On the independence number of random cubic graphs. *Random Structures and Algorithms*, 5:649–664, 1994.

[37] S. L. Hakimi. On the degrees of the vertices of a directed graph. *J. Franklin Inst.*, 279:290–308, 1965.

[38] S. Janson and M. Luczak. A simple solution to the k-core problem. Technical Report 2005:31, Uppsala, 2005.

[39] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. Wiley, 2000.

[40] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

[41] P. T. Johnstone. *Stone Spaces*. Cambridge University Press, 1982.

[42] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, 2002.

[43] R. Karp and M. Sipser. Maximum matchings and sparse random graphs. In *Proc. FOCS*, pages 364–276, 1981.

[44] R. K. Karp. Matchings, cores, and dense subgraphs in sparse random graphs. 1999. Presentation at DIMACS workshop on probabilistic analysis of algorithms for hard problems.

[45] R. M. Karp, M. Luby, and F. M. auf der Heide. Efficient pram simulation on a distributed memory machine. In *Proc. STOC*, pages 318–326, 1992.

[46] J. H. Kim. The poisson cloning model for random graphs, random directed graphs and random k-sat problems. In *Proc. COCOON*, page 2, 2004.

[47] J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. The web as a graph: Measurements, models, and methods. In *Proceedings, International Conference on Combinatorics and Computing*, pages 26–28, 1999.

[48] L. Lovász and V. Sós. Generalized quasirandom graphs. *J. Comb Theory B*, 2007. `http://www.cs.elte.hu/ lovasz/quasirandom.pdf`.

[49] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Comb Theory B*, 96:933–957, 2006.

[50] L. Lu. The diameter of random massive graphs. In *Proc. SODA*, pages 912–921, 2001.

[51] E. Malalla. *Two-way Hashing with Separate Chaining and Linear Probing*. PhD Thesis, School of Computer Science, McGill University, 2004.

[52] B. D. McKay and N. C. Wormald. Asymptotic enumerate by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11:369–382, 1991.

[53] M. Mézard, G. Parisi, and M. Virasoro. *Spin Glass Theory and Beyond*. World Scientific, Singapore, 1988.

[54] M. Molloy. The pure literal rule threshold and cores in random hypergraphs. In *SODA*, pages 665–674, 2004.

[55] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms*, 6:161–179, 1995.

[56] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Comb. Prob. Comput.*, 7:295–305, 1998.

[57] J. R. Munkres. *Topology*. Prentice Hall, 2000.

[58] R. Pagh and F. F. Rodler. Cuckoo hashing. In *Proceedings of the European Sympoisum on Algorithms*, pages 121–133, 2001.

[59] B. Pittel, J. Spencer, and N. Wormald. Sudden emergence of a giant k-core in a random graph. *J. of Combinatorial Theory, Series B*, 67:111–151, 1996.

[60] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.

[61] P. Sanders, S. Egned, and J. Korst. Fast concurrent access to parallel disks. *Algorithmica*, 35(1):21–55, 2003.

[62] P. Sanders, S. Egned, and J. Korst. Efficient hashing with lookups in two memory accesses. In *Proc. SODA*, pages 830–039, 2005.

[63] M. Talagrand. The Parisi formula. *Ann. of Math.*, 162(1):221–263, 2006.

[64] S. Vickers. *Topology Via Logic*. Cambridge University Press, 1989.

[65] N. Wormald. Differential equations for random processes and random graphs. *Annals of Applied Probability*, 5(4):1217–1235, 1995.

Vita

Daniel Fernholz was born in New York City, NY on October 3, 1976, the son of E. Robert and Luisa Fernholz. He graduated from Princeton High School in 1994, and recieved an A.B. in Mathematics, *magna cum laude*, from Harvard University in 1999. From 1999 to 2001, he worked as a consultant and software developer for Trilogy Software, and he began his graduate studies at the University of Texas at Austin in the fall of 2001.

Permanent address: 800 W 3rd St #3336, Austin, Texas 78701

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.